

# Multi-type Galton-Watson processes with affinity-dependent selection applied to antibody affinity maturation

Irene Balelli · Vuk Milišić · Gilles Wainrib

Received: date / Accepted: date

**Abstract** We analyze the interactions between division, mutation and selection in a simplified evolutionary model, assuming that the population observed can be classified into fitness levels. The construction of our mathematical framework is motivated by the modeling of antibody affinity maturation of B-cells in Germinal Centers during an immune response. This is a key process in adaptive immunity leading to the production of high affinity antibodies against a presented antigen. Our aim is to understand how the different biological parameters affect the system's functionality. We identify the existence of an optimal value of the selection rate, able to maximize the number of selected B-cells for a given generation.

**Keywords** Multi-type Galton-Watson process · Germinal center reaction · Affinity-dependent selection · Evolutionary landscapes

**Mathematics Subject Classification (2010)** 60J80 · 60J85 · 60J85

## Contents

1	Introduction . . . . .	2
2	Definitions and modeling assumptions . . . . .	4
3	Results . . . . .	6

I. Balelli · V. Milišić  
Université Paris 13, Institut Galilée, Département de Mathématiques.  
99, avenue Jean-Baptiste Clément 93430 - Villetaneuse - France.  
Tel.: +33-1-49-40-35-39 ; +33-1-49-40-35-91  
Fax: +33-1-49-40-35-68  
E-mail: balelli@math.univ-paris13.fr  
E-mail: milisic@math.univ-paris13.fr

G. Wainrib  
Ecole Normale Supérieure, Département d'Informatique.  
45 rue d'Ulm, 75005 - Paris - France.  
E-mail: gilles.wainrib@ens.fr

4	Extensions of the model . . . . .	23
5	Conclusions and perspectives . . . . .	31

## 1 Introduction

Antibody Affinity Maturation (AAM) takes place in Germinal Centers (GCs), specialized micro-environnements which form in the peripheral lymphoid organs upon infection or immunization [27,9]. GCs are seeded by ten to hundreds distinct B-cells [24], activated after the encounter with an antigen, which initially undergo a phase of intense proliferation [9]. Then, AAM is achieved thanks to multiple rounds of division, Somatic Hypermutation (SHM) of the B-cell receptor proteins, and subsequent selection of B-cells with improved ability of antigen-binding [14]. Those B-cells which successfully complete the GC reaction output as memory B-cells or plasma cells [28,9]. Indirect evidences suggest that only B-cells exceeding a certain threshold of antigen-affinity differentiate into plasma cells [20]. The efficiency of GCs is assured by the contribution of other immune molecules, for instance Follicular Dendritic Cells (FDCs) and follicular helper T-cells (Tfh). Nowadays the key dynamics of GCs are well characterized [14,9,10,24]. Despite this there are still mechanisms which remain unclear, such as the dynamics of clonal competition of B-cells, hence how the selection acts. In recent years a number of mathematical models of the GC reaction appear to investigate these questions, such as [15,30], where the authors have developed agent-based models, mostly analyzed through extensive numerical simulations, or [34] where the authors have established a coarse-grained model, looking for optimal values of *e.g.* the selection strength and the initial B-cell fitness maximizing the affinity improvement.

Our aim in this paper is to contribute to the mathematical foundations of adaptive immunity by introducing and study a simplified evolutionary model inspired by AAM, including division, mutation, affinity-dependent selection and death. We focus on interactions between these mechanisms, identify and analyze the parameters which mostly influence the system functionality, through a rigorous mathematical analysis. This research is motivated by important biotechnological applications. The fundamental understanding of the evolutionary mechanisms involved in AAM have been inspiring many methods for the synthetic production of specific antibodies for drugs, vaccines or cancer immunotherapy [2,13,22]. Indeed, this production process involves the selection of high affinity peptides and requires smart methods to generate an appropriate diversity [8]. Beyond biomedical motivations, the study of this learning process has also given rise in recent years to a new class of bio-inspired algorithms [7,18,25], mainly addressed to solve optimization and learning problems.

We consider a model in which B-cells are classified into  $N + 1$  affinity classes with respect to a presented antigen,  $N$  being an integer big enough to opportunely describe the possible fitness levels of a B-cell with respect to

to a specific antigen [31,33]. A B-cell is able to increase its fitness thanks to SHMs of its receptors: only about 20% of all mutations are estimated to be affinity-affecting mutations [21,23]. By conveniently define a transition probability matrix, we can characterize the probability that a B-cell belonging to a given affinity class pass to another one by mutating its receptors thanks to SHMs. Therefore we define a selection mechanism which acts on B-cells differently depending on their fitness. We mainly focus on a model of positive and negative selection in which B-cells submitted to selection either die or exit the GC center as output cells, according to the strength of their affinity with the antigen. Hence, in this case, no recycling mechanism is taken into account. Nevertheless the framework we set is very easy to manipulate: we can define and study other kinds of affinity-dependent selection mechanisms, and eventually include recycling mechanisms, which have been demonstrated to play an important role in AAM [29]. We demonstrate that independently from the transition probability matrix defining the mutational mechanism and the affinity threshold chosen for positive selection, the optimal selection rate maximizing the number of output cells for the  $t^{\text{th}}$  generation is  $1/t$  (Corollary 2).

From a mathematical point of view, we study a class of multi-types Galton-Watson (GW) processes in which, by considering dead and selected B-cells as two distinct types, we are able to formalize the evolution of a population submitted to an affinity-dependent selection mechanism. To our knowledge, the problem of affinity-dependent selection in GW processes has not been deeply investigated so far.

In Section 2 we define the main model analyzed in this paper. We give as well some definitions that we will use in next sections, such as affinity classes and mutational model. Section 3 contains the main mathematical results. A conveniently use of a multi-type GW process allows to study the evolution of both GC and output cells during time. Proposition 6 collects the formulas which describe the expected size and average affinity of both populations. Moreover, in Section 3.3 we determine the optimal value of the selection rate which maximizes the expected number of selected B-cells at time  $t$ . This value is  $1/t$  independently from all other parameters. We conclude Section 3 with some numerical simulations. In Section 4 we define two possible variants of the model described in previous sections, and show some mathematical results and numerical simulations as well. Section 4 evidences how the mathematical tools used in Section 3 easily apply to define other affinity-dependent selection models. Finally, in Section 5 we discuss our modeling assumptions and give possible extensions and limitations of our mathematical model.

## 2 Definitions and modeling assumptions

This section provides the mathematical framework of this article. Let us suppose that given an antigen target cell  $\bar{\mathbf{x}}$ , all B-cell traits can be divided in exactly  $N + 1$  distinct affinity classes, named 0 to  $N$ .

**Definition 1** Let  $\bar{\mathbf{x}}$  be the antigen target trait. Given a B-cell trait  $\mathbf{x}$ , we denote by  $a_{\bar{\mathbf{x}}}(\mathbf{x})$  the affinity class it belongs to with respect to  $\bar{\mathbf{x}}$ ,  $a_{\bar{\mathbf{x}}}(\mathbf{x}) \in \{0, \dots, N\}$ . The maximal affinity corresponds to the first class, 0, and the minimal one to  $N$ .

**Definition 2** Let  $\mathbf{x}$  be a B-cell trait belonging to the affinity class  $a_{\bar{\mathbf{x}}}(\mathbf{x})$  with respect to  $\bar{\mathbf{x}}$ . We say that its affinity with  $\bar{\mathbf{x}}$  is given by:

$$\text{aff}(\mathbf{x}, \bar{\mathbf{x}}) = N - a_{\bar{\mathbf{x}}}(\mathbf{x})$$

Of course, this is not the only possible choice of affinity. Typically affinity is represented as a Gaussian function [30, 15], having as argument the distance between the B-cell trait and the antigen in the shape space of possible traits. In our model this distance corresponds to the index of the affinity class the B-cell belongs to. Nevertheless the choice of the affinity function does not affect our model.

During the GC reaction B-cells are submitted to random mutations. This implies switches from one affinity class to another with a given probability.

**Definition 3** Let  $(\mathbf{X}_t)_{t \geq 0}$  be a RW on the state-space of B-cell traits describing a pure mutational process of a B-cell during the GC reaction. We denote by  $\mathcal{Q}_N = (q_{ij})_{0 \leq i, j \leq N}$  the transition probability matrix over  $\{0, \dots, N\}$  which gives the probability of passing from an affinity class to another during the introduced mutational model. For all  $0 \leq i, j \leq N$ :

$$q_{ij} = \mathbb{P}(a_{\bar{\mathbf{x}}}(\mathbf{X}_{t+1}) = j \mid a_{\bar{\mathbf{x}}}(\mathbf{X}_t) = i)$$

The main model we study in this paper is defined as follows:

**Definition 4** The process starts with  $z_0 \geq 1$  B-cells entering the GC, belonging to some affinity classes in  $\{0, \dots, N\}$ . In case they are all identical, we denote by  $a_0$  the affinity class they belong to, with respect to the antigen target cell  $\bar{\mathbf{x}}$ . At each time step, each GC B-cell can die with a given rate  $r_d$ . If not, each B-cell can divide with rate  $r_{div}$ : each daughter cell may have a mutated trait, according to the mutational rule allowed. Hence it eventually belongs to a different affinity class than its mother cell. Clearly, it also happens that a B-cell stays in the GC without neither die nor divide. Finally, with rate  $r_s$  each B-cell can be submitted to selection, which is made according to its affinity with  $\bar{\mathbf{x}}$ . A threshold  $\bar{a}_s$  is fixed: if the B-cell belongs to an affinity class with index greater than  $\bar{a}_s$ , the B-cell dies. Otherwise, the B-cell exits the GC pool and reaches the selected pool. Therefore, for any GC B-cell and at any generation, we have:

- $\mathbb{P}(\text{death}) = r_d$
- $\mathbb{P}(\text{division}) = r_{div}$
- $\mathbb{P}(\text{selection}) = r_s$

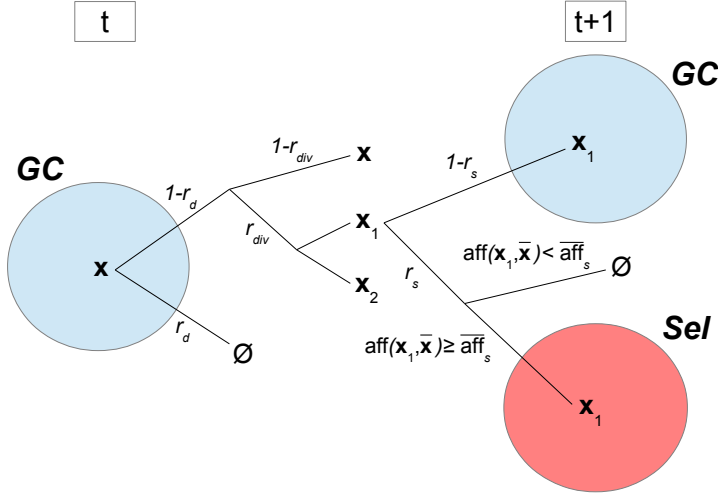


Figure 1: Schematic representation of model described by Definition 4. Here we denote by  $\text{aff}_s$  the fitness corresponding to the affinity class  $\bar{a}_s$ .

The mutation rule reflects the edge set associated to the state-space  $\{0, \dots, N\}$ : this is given by a transition probability matrix.

Once the GC reaction is fully established ( $\sim$  day 7 after immunization), it is polarized into two compartments, named Dark Zone (DZ) and Light Zone (LZ) respectively. The DZ is characterized by densely packed dividing B-cells, while the LZ is less densely populated and contains FDCs and Tfh cells. This is the preferential zone for selection [9]. The transition of B-cells from the DZ to the LZ seems to be determined by a timed cellular program: over a 6 hours period about 50% of DZ B-cells transits to the LZ, where they compete for positive selection signaling [6, 26]. In our simplified mathematical model we do not take into account any spatial factor and in a single time step a GC B-cell can eventually undergo both division (with mutation) and selection. Hence the time unity has to be chosen big enough to take into account both mechanisms.

In previous works [4, 5] we modeled B-cells and antigens as  $N$ -length binary strings, hence their traits correspond to elements of  $\{0, 1\}^N$ . In this context

we characterized affinity thanks to the Hamming distance between B-cell and antigen representing strings.

**Definition 5** For all  $\mathbf{x} = (x_1, \dots, x_N)$ ,  $\mathbf{y} = (y_1, \dots, y_N) \in \{0, 1\}^N$ , their Hamming distance is given by:

$$h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N \delta_i \quad \text{where} \quad \delta_i = \begin{cases} 1 & \text{if } x_i \neq y_i \\ 0 & \text{otherwise} \end{cases}$$

Consequently, in this specific case, the  $i^{\text{th}}$ -affinity class contains B-cells having Hamming distance  $i$  from  $\bar{\mathbf{x}}$  and the fitness is defined as follows:

**Definition 6** For all  $\mathbf{x}_i \in \{0, 1\}^N$ , its affinity with a given vertex  $\bar{\mathbf{x}}$  is given by  $\text{aff}(\mathbf{x}_i, \bar{\mathbf{x}}) := N - h(\mathbf{x}_i, \bar{\mathbf{x}})$ .

While performing numerical simulations (Sections 3.4 and 4.2) we refer to the following transition probability matrix on  $\{0, \dots, N\}$ :

**Definition 7** For all  $i, j \in \{0, \dots, N\}$ :

$$q_{ij} = \mathbb{P}(h(\mathbf{X}_t, \bar{\mathbf{x}}) = j \mid h(\mathbf{X}_t, \bar{\mathbf{x}}) = i) = \begin{cases} i/N & \text{if } j = i - 1 \\ (N - i)/N & \text{if } j = i + 1 \\ 0 & \text{if } |j - i| \neq 1 \end{cases}$$

$\mathcal{Q}_N := (q_{ij})_{0 \leq i, j \leq N}$  is a tridiagonal matrix where the main diagonal consists of zeros.

If we model B-cell traits as vertices of the state-space  $\{0, 1\}^N$ , this corresponds to a model of simple point mutations (see [4] for more details and variants of this basic mutational model on binary strings).

Except for numerical simulations, in this paper we do not restrict to Definitions 5 to 7. All mathematical results obtained in following sections are independent from the hypotheses corresponding to Definitions 5-7. Indeed, in order to define our model we only need to determine  $N + 1$  distinct affinity classes of B-cell traits with respect to a presented antigen and the probabilities that a GC B-cell passes from a given affinity class to another one thanks to somatic hypermutations during the GC reaction.

### 3 Results

In this Section we formalize mathematically the model introduced above. This enables the estimation of various qualitative and quantitative measures of the GC evolution and of the selected pool as well. In Section 3.1 we show that a simple GW process describes the evolution of the size of the GC and determine a condition for its extinction. In order to do this we do not need to know the mutational model. Nevertheless, if we want to understand deeply the whole

reaction we need to consider a  $(N + 3)$ -type GW process, which we introduce in Section 3.2. Therefore we determine explicitly other quantities, such as the average affinity in the GC and the selected pool, or the evolution of the size of the latter. We conclude this section by numerical simulations (Section 3.4).

### 3.1 Evolution of the GC size

The aim of this section is to estimate the evolution of the GC size and its extinction probability. In order to do so we define a simple GW process, with respect to the parameters  $r_d$ ,  $r_{div}$  and  $r_s$ . Indeed, each B-cell submitted to selection exits the GC pool, independently from its affinity with  $\bar{x}$ . Hence we apply some classical results about generating functions and GW processes (see [11], chapter I). We collect these results for our specific case in Theorem 1. Corollary 1 gives explicitly the expected size of the GC at time  $t$  and conditions for the extinction of the GC.

**Definition 8** Let  $Z_t^{(z_0)}$ ,  $t \geq 0$  be the random variable describing the GC-population size at time  $t$ , starting from  $z_0 \geq 1$  initial B-cells.  $(Z_t^{(z_0)})_{t \in \mathbb{N}}$  is a MC (as each cell behaves independently from the others and from the previous generations) on  $\{0, 1, 2, \dots\}$ .

If  $z_0 = 1$  and there is no confusion, we denote  $Z_t := Z_t^{(1)}$ . By Definition 8,  $Z_1$  corresponds to the number of cells in the GC at the first generation, starting from a single seed cell. Thanks to Definition 4 one can claim that  $Z_1 \in \{0, 1, 2\}$ , with the following probabilities:

$$\begin{cases} p_0 := \mathbb{P}(Z_1 = 0) = r_d + (1 - r_d)r_s(1 - r_{div} + r_{div}r_s) \\ p_1 := \mathbb{P}(Z_1 = 1) = (1 - r_d)(1 - r_s)(1 - r_{div} + 2r_{div}r_s) \\ p_2 := \mathbb{P}(Z_1 = 2) = r_{div}(1 - r_d)(1 - r_s)^2 \end{cases} \quad (1)$$

As far as next generations are concerned, conditioning to  $Z_t = k$ ,  $Z_{t+1}$  is distributed as the sum of  $k$  independent copies of  $Z_1$ , i.e.  $\mathbb{P}(Z_{t+1} = k' | Z_t = k) = \mathbb{P}\left(\sum_{i=1}^k Z_1 = k'\right)$ .

**Definition 9** Let  $X$  be an integer valued random variable (rv),  $p_k := \mathbb{P}(X = k)$  for all  $k \geq 0$ . Its probability generating function (pgf) is given by:

$$F_X(s) = \sum_{k=0}^{+\infty} p_k s^k$$

$F_X$  is a convex monotonically increasing function over  $[0, 1]$ , and  $F_X(1) = 1$ . If  $p_0 \neq 0$  and  $p_0 + p_1 < 1$  then  $F$  is a strictly increasing function.

**Definition 10** Given  $F$ , the pgf of a rv  $X$ , the iterates of  $F$  are given by:

$$\begin{aligned} F_0(s) &= s \\ F_1(s) &= F(s) \\ F_t(s) &= F(F_{t-1}(s)) \text{ for } t \geq 2 \end{aligned}$$

**Proposition 1**

- (i) If  $\mathbb{E}(X)$  exists (respectively  $\mathbb{V}(X)$ ), then  $\mathbb{E}(X) = F'_X(1)$  (respectively  $\mathbb{V}(X) = F''_X(1) - (\mathbb{E}(X))^2 + \mathbb{E}(X)$ ).
- (ii) If  $X$  and  $Y$  are two integer valued independent rvs, then  $X + Y$  is still an integer valued rv and its pgf is given by  $F_{X+Y} = F_X F_Y$ .

The pgf for  $Z_1$  is given by:

$$\begin{aligned} F(s) &= p_0 + p_1 s + p_2 s^2 \\ &= r_d + (1 - r_d)r_s(1 - r_{div} + r_{div}r_s) \\ &\quad + (1 - r_d)(1 - r_s)(1 - r_{div} + 2r_{div}r_s)s + r_{div}(1 - r_d)(1 - r_s)^2 s^2 \end{aligned} \quad (2)$$

**Definition 11** We denote by  $\eta$  the extinction probability of the process  $(Z_t)_{t \in \mathbb{N}}$ :

$$\eta := \lim_{t \rightarrow \infty} F_t(0)$$

**Theorem 1**

- (i) The pgf of  $Z_t^{(z_0)}$ ,  $t \in \mathbb{N}$ , which represents the population size of the  $t^{\text{th}}$ -generation starting from  $z_0 \geq 1$  seed cells, is  $F_t^{(z_0)} = (F_t)^{z_0}$ ,  $F_t$  being the  $t^{\text{th}}$ -iterate of  $F$  (Equation (2)).
- (ii) The expected size of the GC at time  $t$  and starting from  $z_0$  B-cells is given by:

$$\mathbb{E}(Z_t^{(z_0)}) = (\mathbb{E}(Z_t))^{z_0} = \left( (\mathbb{E}(Z_1))^t \right)^{z_0}, \quad (3)$$

- (iii)  $\eta$  is the smallest fixed point of the generating function  $F$ , i.e.  $\eta$  is the smallest  $s$  s.t.  $F(s) = s$ .
- (iv) If  $\mathbb{E}(Z_1) =: m$  is finite, then:
- if  $m \leq 1$  then  $F$  has only 1 as fixed point and consequently  $\eta = 1$ ;
  - if  $m > 1$  then  $F$  has exactly a fixed point on  $[0, 1[$  and then  $\eta < 1$ .
- (v) Denoted by  $\eta_{z_0}$  the probability of extinction of  $(Z_t^{(z_0)})$ , one has:

$$\eta_{z_0} = \eta^{z_0}$$

where  $\eta$  is given by (iii).

By applying Theorem 1 and Equation (1) above, one can prove:

**Corollary 1**

- (i) The expected size of the GC at time  $t$  and starting from  $z_0$  initial B-cells is given by:

$$\mathbb{E}(Z_t^{(z_0)}) = ((1 - r_d)(1 + r_{div})(1 - r_s))^{tz_0} \quad (4)$$

- (ii) Denoted by  $\eta_{z_0}$  the extinction probability of the GC population starting from  $z_0$  initial B-cells, one has :

- if  $r_s \geq 1 - \frac{1}{(1 - r_d)(1 + r_{div})}$ , then  $\eta_{z_0} = 1$
- otherwise  $\eta_{z_0} = \eta^{z_0} < 1$ ,  $\eta$  being the smallest fixed point of (2)



Then in particular the process is subcritical or supercritical independently from  $z_0$ . In the supercritical case, increasing the number of B-cells at the beginning of the process makes the probability of extinction decrease. More precisely, in the case  $\eta < 1$ , then  $\eta_{z_0} \rightarrow 0$  if  $z_0 \rightarrow \infty$ .

This section shows that a classical use of a simple GW process enables to understand quantitatively the GC growth. Moreover, Corollary 1 (ii) gives a condition over the main parameters for the extinction of the GC : if the selection pressure is too high, with probability 1 the GC size goes to 0, independently from the initial number of seed cells. Intuitively, a too high selection pressure prevents those B-cells with bad affinity to improve their fitness undergoing further rounds of mutation and division. Most B-cells will be rapidly submitted to selection, hence either exit the GC as output cells or die by apoptosis if they fail to receive positive selection signals [14].

### 3.2 Evolution of the size and fitness of GC and selected pools

The GW process defined in the previous Section only describes the size of the GC. Indeed, we are not able to say anything about the average fitness of GC clones, or the expected number of selected B-cells, or their average affinity. Hence, we need to consider a more complex model and take into account the parameter  $\bar{a}_s$  and the transition probability matrix characterizing the mutational rule. We introduce a multi-type GW Process (see for instance [3], chapter V).

**Definition 12** Let  $\mathbf{Z}_t^{(i)} = (Z_{t,0}^{(i)}, \dots, Z_{t,N+2}^{(i)})$ ,  $t \geq 0$  be a MC where for all  $0 \leq j \leq N$ ,  $Z_{t,j}^{(i)}$  describes the number of GC B-cells belonging to the  $j^{\text{th}}$ -affinity class with respect to  $\bar{\mathbf{x}}$ ,  $Z_{t,N+1}^{(i)}$  the number of selected B-cells and  $Z_{t,N+2}^{(i)}$  the number of dead B-cells at generation  $t$ , when the process is initiated in state  $\mathbf{i} = (i_0, \dots, i_N, 0, 0)$ .

For all  $j \in \{0, \dots, N+2\}$  the generating function gives the number of offsprings of each type that a type  $j$  particle can produce. It is defined as follows:

$$f^{(j)}(s_0, \dots, s_{N+2}) = \sum_{k_0, \dots, k_{N+2} \geq 0} p^{(j)}(k_0, \dots, k_{N+2}) s_0^{k_0} \dots s_{N+2}^{k_{N+2}}, \quad (5)$$

$$0 \leq s_\alpha \leq 1 \text{ for all } \alpha \in \{0, \dots, N+2\}$$

where  $p^{(j)}(k_0, \dots, k_{N+2})$  is the probability that a type  $j$  cell produces  $k_0$  cells of type 0,  $k_1$  of type 1, ...,  $k_{N+2}$  of type  $N+2$  for the next generation.

We denote:

- $\mathbf{p}(\mathbf{k}) = (p^{(0)}(\mathbf{k}), \dots, p^{(N+2)}(\mathbf{k}))$ , for  $\mathbf{k} = (k_0, \dots, k_{N+2}) \in \mathbb{Z}_+^{N+3}$
- $\mathbf{f}(\mathbf{s}) = (f^{(1)}(\mathbf{s}), \dots, f^{(N+1)}(\mathbf{s}))$ , for  $\mathbf{s} = (s_0, \dots, s_{N+2}) \in \mathcal{C}^{N+3}$  where  $\mathcal{C}^{N+3} := \{\mathbf{x} \in \mathbb{R}^{N+3} \mid 0 \leq x_\alpha \leq 1, \alpha \in \{0, \dots, N+2\}\}$

Then the probability generating function of  $\mathbf{Z}_1$  is given by:

$$\mathbf{f}(\mathbf{s}) = \sum_{\mathbf{k} \in \mathbb{Z}_+^{N+3}} \mathbf{p}(\mathbf{k}) \mathbf{s}^{\mathbf{k}}, \mathbf{s} \in \mathcal{C}^{N+3} \quad (6)$$

Again, the generating function of  $\mathbf{Z}_t$ ,  $\mathbf{f}_t(\mathbf{s})$ , is obtained as the  $t^{\text{th}}$ -iterate of  $\mathbf{f}$ , and it holds true that:

$$\mathbf{f}_{t+r}(\mathbf{s}) = \mathbf{f}_t[\mathbf{f}_r(\mathbf{s})], \mathbf{s} \in \mathcal{C}^{N+3}.$$

Let  $m_{ij} := \mathbb{E}[Z_{1,j}^{(i)}]$  the expected number of offspring of type  $j$  of a cell of type  $i$  in one generation. We collect all  $m_{ij}$  in a matrix,  $\mathcal{M} = (m_{ij})_{0 \leq i, j \leq N+2}$ . We have:

$$m_{ij} = \frac{\partial f^{(i)}}{\partial s_j}(\mathbf{1})$$

and:

$$\mathbb{E}[Z_{t,j}^{(i)}] = \frac{\partial f_t^{(i)}}{\partial s_j}(\mathbf{1}) \quad (7)$$

Finally:

$$\mathbb{E}[\mathbf{Z}_t^{(i)}] = \mathbf{i} \mathcal{M}^t \quad (8)$$

We can now explicitly give the elements of  $\mathcal{M}$  by using matrix  $\mathcal{Q}_N$  given by Definition 3.

**Proposition 2**  $\mathcal{M}$  is a  $(N+3) \times (N+3)$  matrix, which we can define as a block matrix in the following way:

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\mathbf{0}_{2 \times (N+1)}$  is a  $2 \times (N+1)$  matrix with all entries 0;
- $\mathcal{I}_n$  is the identity matrix of size  $n$ ;
- $\mathcal{M}_1 = 2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$
- $\mathcal{M}_2 = (m_{2,ij})$  is a  $(N+1) \times 2$  matrix where for all  $i \in \{0, \dots, N\}$ :
  - if  $i \leq \bar{a}_s$ :

$$m_{2,i1} = (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=0}^{\bar{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + 2(1-r_d)r_{div}r_s \sum_{j=\bar{a}_s+1}^N q_{ij}$$

- if  $i > \bar{a}_s$ :

$$m_{2,i1} = 2(1-r_d)r_{div}r_s \sum_{j=0}^{\bar{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=\bar{a}_s+1}^N q_{ij}$$

*Proof* It suffices to compute  $f^{(i)}(\mathbf{s})$  for  $i = 0, \dots, N+2$ , which depend on  $r_d$ ,  $r_{div}$ ,  $r_s$  and the elements of  $\mathcal{Q}_N$ . First, the elements of the  $(N+2)^{\text{th}}$  and  $(N+3)^{\text{th}}$ -lines are obviously determined : all selected (resp. dead) cells remain selected (resp. dead) for next generations, as they can not give rise to any other offspring type cell (we do not take into account here any type of recycling mechanism). Let  $i \in \{0, \dots, N\}$  be a fixed index: we evaluate  $m_{ij}$  for all  $j \in \{0, \dots, N+2\}$ . The first step is to determine the value of  $p^{(i)}(\mathbf{k})$  for  $\mathbf{k} = (k_0, \dots, k_{N+2}) \in \mathbb{Z}_+^{N+3}$ . There exists only a few cases in which  $p^{(i)}(\mathbf{k}) \neq 0$ , which can be explicitly evaluated:

$$\begin{aligned}
- p^{(i)}(0, \dots, 0, 1) &= \begin{cases} r_d & \text{if } i \leq \bar{a}_s \\ r_d + (1 - r_d)(1 - r_{div})r_s & \text{otherwise} \end{cases} \\
- p^{(i)}(0, \dots, 0, 1, 0) &= \begin{cases} (1 - r_d)(1 - r_{div})r_s & \text{if } i \leq \bar{a}_s \\ 0 & \text{otherwise} \end{cases} \\
- p^{(i)}(0, \dots, 0, 1, 0, \dots, 0, 0) &= (1 - r_d)(1 - r_{div})(1 - r_s) \\
- p^{(i)}(0, \dots, 0, 2) &= (1 - r_d)r_{div}r_s^2 \sum_{j_1=\bar{a}_s+1}^N q_{ij_1} \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} \\
- p^{(i)}(0, \dots, 0, 2, 0) &= (1 - r_d)r_{div}r_s^2 \sum_{j_1=0}^{\bar{a}_s} q_{ij_1} \sum_{j_2=0}^{\bar{a}_s} q_{ij_2} \\
- p^{(i)}(0, \dots, 0, 1, 1) &= 2(1 - r_d)r_{div}r_s^2 \sum_{j_1=0}^{\bar{a}_s} q_{ij_1} \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} \\
- \text{For all } j_1 < j_2 \in \{0, \dots, N\}: \\
- p^{(i)}(0, \dots, 0, 2, 0, \dots, 0, 0) &= (1 - r_d)r_{div}(1 - r_s)^2 q_{ij_1}^2 \\
- p^{(i)}(0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0, 0) &= 2(1 - r_d)r_{div}(1 - r_s)^2 q_{ij_1} q_{ij_2} \\
- p^{(i)}(0, \dots, 0, 1, 0, \dots, 0, 1) &= 2(1 - r_d)r_{div}r_s(1 - r_s) q_{ij_1} \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} \\
- p^{(i)}(0, \dots, 0, 1, 0, \dots, 0, 1, 0) &= 2(1 - r_d)r_{div}r_s(1 - r_s) q_{ij_1} \sum_{j_2=0}^{\bar{a}_s} q_{ij_2} \\
- p^{(i)}(\mathbf{k}) &= 0 \text{ otherwise}
\end{aligned}$$

We can therefore evaluate  $f^{(i)}(\mathbf{s})$ , with  $\mathbf{s} = (s_0, \dots, s_{N+2}) \in \mathcal{C}^{N+3}$ .

For all  $i \leq \bar{a}_s$ :

$$\begin{aligned}
f^{(i)}(\mathbf{s}) &= r_d s_{N+2} + (1-r_d)(1-r_{div})r_s s_{N+1} + (1-r_d)(1-r_{div})(1-r_s)s_i \\
&+ (1-r_d)r_{div}r_s^2 \left( \sum_{j_1=\bar{a}_s+1}^N q_{ij_1} \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} s_{N+2}^2 \right. \\
&\left. + \sum_{j_1=0}^{\bar{a}_s} q_{ij_1} \sum_{j_2=0}^{\bar{a}_s} q_{ij_2} s_{N+1}^2 + 2 \sum_{j_1=0}^{\bar{a}_s} q_{ij_1} \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} s_{N+1} s_{N+2} \right) \\
&+ (1-r_d)r_{div}(1-r_s)^2 \left( \sum_{j_1=0}^N q_{ij_1}^2 s_{j_1}^2 + 2 \sum_{j_1=0}^N q_{ij_1} \sum_{j_2 < j_1=0}^N q_{ij_2} s_{j_1} s_{j_2} \right) \\
&+ 2(1-r_d)r_{div}r_s(1-r_s) \sum_{j_1=0}^N q_{ij_1} \left( \sum_{j_2=\bar{a}_s+1}^N q_{ij_2} s_{N+2} + \sum_{j_2=0}^{\bar{a}_s} q_{ij_2} s_{N+1} \right) s_{j_1}
\end{aligned} \tag{9}$$

If  $i > \bar{a}_s$  then  $f^{(i)}(\mathbf{s})$  is the same except for the first line, which becomes:

$$(r_d + (1-r_d)(1-r_{div})r_s)s_{N+2} + (1-r_d)(1-r_{div})(1-r_s)s_i$$

The values of each  $m_{ij}$  is now obtained by evaluating all partial derivatives of  $f^{(i)}(\mathbf{s})$  in **1**, keeping in mind that for all  $i \in \{0, \dots, N\}$ ,  $\sum_{j=0}^N q_{ij} = 1$ .  $\square$

*Example 1* One can give explicitly the form of matrix  $\mathcal{M}_2$  corresponding to the mutational model defined in Definition 7:

$$\mathcal{M}_2 = \begin{matrix} 0 \\ \vdots \\ \bar{a}_s - 1 \\ \bar{a}_s \\ \bar{a}_s + 1 \\ \bar{a}_s + 2 \\ \vdots \\ N \end{matrix} \begin{pmatrix} \alpha & r_d \\ \vdots & \vdots \\ \alpha & r_d \\ \alpha - \beta + \beta \frac{\bar{a}_s}{N} & r_d + \beta \frac{N - \bar{a}_s}{N} \\ \beta \frac{\bar{a}_s + 1}{N} & r_d + \alpha - \beta + \beta \frac{N - (\bar{a}_s + 1)}{N} \\ 0 & r_d + \alpha \\ \vdots & \vdots \\ 0 & r_d + \alpha \end{pmatrix},$$

where:

$$\begin{aligned}
- \alpha &:= (1-r_d)(1+r_{div})r_s \\
- \beta &:= 2(1-r_d)r_{div}r_s
\end{aligned}$$

*Remark 1* Independently from the introduced mutational model,  $\alpha + r_d$  corresponds to the expected number of selected or dead B-cells that each GC B-cell can produce in a single time step.

Of course in the multi-type context we recover again results from Section 3.1. For this sake, let us recall some results about the extinction probability for multi-type GW processes [3].

**Definition 13** Let  $q^{(i)}$  be the probability of eventual extinction of the process, when it starts from a single type  $i$  cell. As above bold symbols denote vectors i.e.  $\mathbf{q} := (q^{(0)}, \dots, q^{(N+2)}) \geq 0$ .

**Definition 14** We say that  $(\mathbf{Z}_t)$  is singular if each particle has exactly one offspring, which implies that the branching process becomes a simple MC.

**Definition 15** Matrix  $\mathcal{M}$  is said to be strictly positive if it has non-negative entries and there exists a  $t$  s.t.  $(\mathcal{M}^t)_{ij} > 0$  for all  $i, j$ .  $(\mathbf{Z}_t)$  is positive regular iff  $\mathcal{M}$  is strictly positive.

*Notation 1* Let  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ . We say that  $\mathbf{u} \leq \mathbf{v}$  if  $u_i \leq v_i$  for all  $i \in \{1, \dots, n\}$ . Moreover, we say that  $\mathbf{u} < \mathbf{v}$  if  $\mathbf{u} \leq \mathbf{v}$  and there exists at least an index  $j$  s.t.  $u_j < v_j$ .

**Theorem 2** Let  $(\mathbf{Z}_t)$  be non singular and strictly positive. Let  $\rho$  be the maximum eigenvalue of  $\mathcal{M}$ . Then the following three results holds:

1. If  $\rho < 1$  (subcritical case) or  $\rho = 1$  (critical case) then  $\mathbf{q} = \mathbf{1}$ . Otherwise, if  $\rho > 1$  (supercritical case), then  $\mathbf{q} < \mathbf{1}$ .
2.  $\lim_{t \rightarrow \infty} \mathbf{f}_t(\mathbf{s}) = \mathbf{q}$ , for all  $\mathbf{s} \in \mathcal{C}^{N+3}$ .
3.  $\mathbf{q}$  is the only solution of  $\mathbf{f}(\mathbf{s}) = \mathbf{s}$  in  $\mathcal{C}^{N+3}$ .

The spectra of matrix  $\mathcal{M}$  defined in Definition 2 is obtained as follows:

**Proposition 3** Let  $\mathcal{M}$  be defined as a block matrix as in 2. Let  $\lambda_{\mathcal{M},i}$  be its  $i^{\text{th}}$ -eigenvalue. The spectra of  $\mathcal{M}$  is given by:

- For all  $i \in \{0, \dots, N\}$ ,  $\lambda_{\mathcal{M},i} = (1 - r_d)(1 - r_s)(1 + r_{div}(2\lambda_i - 1))$ , where  $\lambda_i$  is the  $i^{\text{th}}$ -eigenvalue of matrix  $\mathcal{Q}_N$ .
- whereas  $\lambda_{\mathcal{M},N+1} = 1$  with multiplicity 2.

*Proof* As  $\mathcal{M}$  is a block matrix with the lower left block composed of zeros, then  $\text{Spec}(\mathcal{M}) = \text{Spec}(\mathcal{M}_1) \cup \text{Spec}(\mathcal{I}_2)$ . The result follows.  $\square$

Therefore we obtain the same condition as in Proposition 1 for the extinction probability in the GC:

**Proposition 4** Let  $\mathbf{q}$  be the extinction probability for the process  $(\mathbf{Z}_t)$  defined in Definition 20 and restricted to the first  $N+1$  components (i.e. we refer only to matrix  $\mathcal{M}_1$ , which defines the expectations of GC B-cells). Therefore:

- if  $r_s \geq 1 - \frac{1}{(1 - r_d)(1 + r_{div})}$ , then  $\mathbf{q} = \mathbf{1}$
- otherwise  $\mathbf{q} < \mathbf{1}$  is the smallest fixed point of  $\mathbf{f}(\mathbf{s})$  in  $\mathcal{C}^{N+3}$ .

*Proof*  $\mathcal{Q}_N$  is a stochastic matrix, therefore its largest eigenvalue is 1. The corresponding eigenvalue of matrix  $\mathcal{M}_1$  is:  $\lambda_{\mathcal{M}_1,1} = (1-r_d)(1-r_s)(1+r_{div})$ . The proposition is proved by observing that  $\lambda_{\mathcal{M}_1,1} \leq 1 \Leftrightarrow r_s \geq 1 - \frac{1}{(1-r_d)(1+r_{div})}$  and applying Theorem 2 (note that  $\mathcal{M}_1$  is positive regular: this is not the case for matrix  $\mathcal{M}$ ).  $\square$

In order to determine the expected number of selected cells at a given time  $t$ , we need to introduce another multi-type GW process.

**Definition 16** Let  $\tilde{\mathbf{Z}}_t^{(i)} = (\tilde{Z}_{t,0}^{(i)}, \dots, \tilde{Z}_{t,N+2}^{(i)})$ ,  $t \geq 0$  be a MC where for all  $0 \leq j \leq N$ ,  $\tilde{Z}_{t,j}^{(i)}$  describes the number of GC B-cells belonging to the  $j^{\text{th}}$ -affinity class with respect to  $\bar{\mathbf{x}}$ ,  $\tilde{Z}_{t,N+1}^{(i)}$  the number of selected B-cells and  $\tilde{Z}_{t,N+2}^{(i)}$  the number of dead B-cells at generation  $t$ , when the process is initiated in state  $\mathbf{i} = (i_0, \dots, i_N, 0, 0)$  and before the selection mechanism is performed for the  $t^{\text{th}}$ -generation.

Proceeding as we did for  $\mathbf{Z}_t^{(i)}$ , we can determine matrix  $\tilde{\mathcal{M}}$  whose elements are  $\tilde{m}_{ij} := \mathbb{E}[\tilde{Z}_{1,j}^{(i)}]$  for all  $i, j \in \{0, \dots, N+2\}$ .

**Proposition 5**  $\tilde{\mathcal{M}}$  is a  $(N+3) \times (N+3)$  matrix, which only depends on matrix  $\mathcal{Q}_N$ ,  $r_d$  and  $r_{div}$  and can be defined as a block matrix as follows:

$$\tilde{\mathcal{M}} = \begin{pmatrix} \tilde{\mathcal{M}}_1 & \tilde{\mathcal{M}}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\tilde{\mathcal{M}}_1 = 2(1-r_d)r_{div}\mathcal{Q}_N + (1-r_d)(1-r_{div})\mathcal{I}_{N+1}$
- $\tilde{\mathcal{M}}_2 = (\mathbf{0}_{N+1}, r_d \cdot \mathbf{1}_{N+1})$ , where  $\mathbf{0}_{N+1}$  (resp.  $\mathbf{1}_{N+1}$ ) is a  $(N+1)$ -column vector whose elements are all 0 (resp. 1).

*Notation 2* Let  $S_t$ ,  $t \geq 0$  be the random variable describing the number of selected B-cells at time  $t$ . By hypothesis  $S_0 = 0$ .  $(S_t)_{t \in \mathbb{N}}$  is a MC on  $\{0, 1, 2, \dots\}$ .

We can therefore prove the following results:

**Proposition 6** Let  $\mathbf{i}$  be the initial state.

- The expected size of the GC at time  $t$  is given by:

$$\sum_{k=0}^N (\mathbf{i}\mathcal{M}^t)_k \tag{10}$$

- The average affinity in GC at time  $t$  is given by:

$$\frac{\sum_{k=0}^N (N-k)(\mathbf{i}\mathcal{M}^t)_k}{\sum_{k=0}^N (\mathbf{i}\mathcal{M}^t)_k} \tag{11}$$

- The expected number of selected B-cells at time  $t$  is given by:

$$\mathbb{E}(S_t) = r_s \sum_{k=0}^{\bar{a}_s} \left( \mathbf{iM}^{t-1} \widetilde{\mathcal{M}} \right)_k \quad (12)$$

- The expected number of selected B-cells produced until time  $t$  is given by:

$$\mathbb{E} \left[ \sum_{n=0}^t S_n \right] = \mathbb{E} \left[ \left( \mathbf{Z}_t^{(\mathbf{i})} \right)_{N+2} \right] = (\mathbf{iM}^t)_{N+2} \quad (13)$$

- The average affinity of selected B-cells at time  $t$  is given by:

$$\frac{\sum_{k=0}^{\bar{a}_s} (N-k) \left( \mathbf{iM}^{t-1} \widetilde{\mathcal{M}} \right)_k}{\sum_{k=0}^{\bar{a}_s} \left( \mathbf{iM}^{t-1} \widetilde{\mathcal{M}} \right)_k} \quad (14)$$

- The average affinity of selected B-cells until time  $t$  is given by:

$$\frac{r_s \sum_{s=1}^t \sum_{k=0}^{\bar{a}_s} (N-k) \left( \mathbf{iM}^{s-1} \widetilde{\mathcal{M}} \right)_k}{(\mathbf{iM}^t)_{N+2}} \quad (15)$$

*Proof* Equations (10), (11) and (13) are a direct application of what stated in Equation (8). In order to prove Equation (13) we have to observe that, as we already commented in the case of a single-type GW process (Equation (23)) :

$$\mathbb{E} \left[ \widetilde{\mathbf{Z}}_t^{(\mathbf{i})} \right] = \mathbf{iM}^{t-1} \widetilde{\mathcal{M}} \quad (16)$$

Moreover, we have to remark that the expected number of selected B-cells at time  $t$  is obtained from the expected number of B-cells in GC at time  $t$  (before the selection mechanism is performed) having fitness good enough to be positive selected. This is clearly given by  $\sum_{k=0}^{\bar{a}_s} \left( \mathbf{iM}^{t-1} \widetilde{\mathcal{M}} \right)_k$ , thanks to (16). We have just to multiply this expectation for the probability that each of these B-cells is submitted to mutation, *i.e.*  $r_s$ . Finally, results about the average affinity in both the GC and the selected pool (Equations (11), (14) and (15)) are obtained from the previous ones by multiplying the number of individuals belonging to the same class by their fitness (Definition 2), and dividing by the total number of individuals in the considered pool. The definition of affinity as function of the affinity classes, determines Equations (11), (14) and (15).  $\square$

The expected size of the GC at time  $t$  can be obtained applying a simple GW process (Section 3.1) and is given by (4). It is possible to prove the same result starting from the  $(N+3)$ -type GW process (10). For the sake of simplicity, let us suppose that the process starts from a single B-cell belonging

to the affinity class  $a_0 = i$  with respect to the target trait. We do not need to specify the transition probability matrix used to define the mutational model allowed.

We can easily prove by iteration that:

$$\mathcal{M}^t = \begin{pmatrix} \mathcal{M}_1^t & \sum_{k=0}^{t-1} \mathcal{M}_1^k \mathcal{M}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix} \quad (17)$$

Therefore we can claim that  $(\mathbf{i}\mathcal{M}^t)_k$  corresponds to the  $k^{\text{th}}$ -component of the  $i^{\text{th}}$ -row of matrix  $\mathcal{M}_1^t = (2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1})^t$ , where  $\mathcal{Q}_N$  is a stochastic matrix. Matrices  $\mathcal{A} := 2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N$  and  $\mathcal{B} := (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$  clearly commute, therefore we write:

$$(\mathcal{A} + \mathcal{B})^t = \sum_{j=0}^t C_t^j \mathcal{A}^{t-j} \mathcal{B}^j \quad (18)$$

For all  $j$ ,  $0 \leq j \leq t$ :

$$\begin{aligned} \mathcal{A}^{t-j} \mathcal{B}^j &= 2^{t-j} (1-r_d)^{t-j} r_{div}^{t-j} (1-r_s)^{t-j} (1-r_d)^j (1-r_{div})^j (1-r_s)^j \mathcal{Q}_N^{t-j} \\ &= (1-r_d)^t (1-r_s)^t (2r_{div})^{t-j} (1-r_{div})^j \mathcal{Q}_N^{t-j} \end{aligned}$$

Hence:

$$(\mathcal{A} + \mathcal{B})^t = (1-r_d)^t (1-r_s)^t \sum_{j=0}^t C_t^j (2r_{div})^{t-j} (1-r_{div})^j \mathcal{Q}_N^{t-j}$$

And consequently:

$$\begin{aligned} \sum_{k=0}^N (\mathbf{i}\mathcal{M}^t)_k &= \sum_{k=0}^N \left( \mathbf{i}(\mathcal{A} + \mathcal{B})^t \right)_k \\ &= (1-r_d)^t (1-r_s)^t \sum_{j=0}^t C_t^j (2r_{div})^{t-j} (1-r_{div})^j \sum_{k=0}^N \left( \mathbf{i}\mathcal{Q}_N^{t-j} \right)_k \end{aligned}$$

Since  $\mathcal{Q}_N$  is a stochastic matrix, for all  $n$ ,  $\mathcal{Q}_N^n$  is still a stochastic matrix, *i.e.* the entries of each row of  $\mathcal{Q}_N^n$  sum to 1. Therefore:

$$\begin{aligned} \sum_{k=0}^N (\mathbf{i}\mathcal{M}^t)_k &= (1-r_d)^t (1-r_s)^t \sum_{j=0}^t C_t^j (2r_{div})^{t-j} (1-r_{div})^j \\ &= (1-r_d)^t (1-r_s)^t (2r_{div} + 1 - r_{div})^t = (1-r_d)^t (1-r_s)^t (1+r_{div})^t, \end{aligned}$$

as stated by Equation (4) for  $z_0 = 1$ . This result can be easily generalized to the case of  $z_0 \geq 1$  initial B-cells.



### 3.3 $r_s$ maximizing the expectation of selected B-cells at time $t$

What is the behavior of the expected number of selected B-cells as a function of the model parameters ? In particular, is there an optimal value of the selection rate which maximizes this number ? In this section we show that, indeed, the answer is positive.

To do so we detail hereafter the computation of  $\mathbb{E}(S_t)$ , given by Proposition 6.

Let us suppose, for the sake of simplicity, that  $\mathcal{Q}_N$  is diagonalizable:

$$\mathcal{Q}_N = R\Lambda_N L, \quad (19)$$

where  $\Lambda_N = \text{diag}(\lambda_0, \dots, \lambda_N)$ , and  $R = (r_{ij})$  (resp.  $L = (l_{ij})$ ) is the transition matrix whose rows (resp. lines) contain the right (resp. left) eigenvectors of  $\mathcal{Q}_N$ , corresponding to  $\lambda_0, \dots, \lambda_N$ . This is the case, for example, if we consider the mutational model given by Definition 7. Moreover, in this specific case, the  $N+1$  distinct eigenvalues of  $\mathcal{Q}_N$  are known explicitly [4]:

$$\lambda_0 = 1 \leq 1 - \frac{1}{N} \leq 1 - \frac{2}{N} \leq \dots \leq -1 + \frac{2}{N} \leq -1 + \frac{1}{N} \leq -1 = \lambda_N$$

It follows from (17) and (19) that for all  $t \geq 1$ ,  $\mathcal{M}^t$  can be written as:

$$\mathcal{M}^t = \begin{pmatrix} RD^t L & \left( R \sum_{k=0}^{t-1} D^k L \right) \mathcal{M}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}, \quad (20)$$

where  $D = 2(1-r_d)r_{div}(1-r_s)\Lambda_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$  is a diagonal matrix. We obtain its expression thanks to Proposition 2.

Moreover, by Proposition 5 and Equation (19) we have:

$$\widetilde{\mathcal{M}} = \begin{pmatrix} R\widetilde{D}L & \widetilde{\mathcal{M}}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}, \quad (21)$$

where  $\widetilde{D} = 2(1-r_d)r_{div}\Lambda_N + (1-r_d)(1-r_{div})\mathcal{I}_{N+1}$  is a diagonal matrix.

**Proposition 7** *Let us suppose that at time  $t = 0$  there is a single B-cell entering the GC belonging to the  $i^{\text{th}}$ -affinity class with respect to the target cell. Moreover, let us suppose that  $\mathcal{Q}_N = R\Lambda_N L$ . For all  $t \geq 1$ , the expected number of selected B-cells at time  $t$ , is:*

$$\mathbb{E}(S_t) = r_s(1-r_s)^{t-1}(1-r_d)^t \sum_{\ell=0}^N (2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{\bar{a}_s} r_{i\ell} l_{\ell k},$$

*Proof* Proposition 6 claims:

$$\mathbb{E}(S_t) = r_s \sum_{k=0}^{\bar{a}_s} \left( \mathbf{i} \mathcal{M}^{t-1} \widetilde{\mathcal{M}} \right)_k$$

We have to explicitly write  $\left( \mathbf{i} \mathcal{M}^{t-1} \widetilde{\mathcal{M}} \right)_k$ . From Equations (20) and (21):

$$\mathcal{M}^{t-1} \widetilde{\mathcal{M}} = \begin{pmatrix} RD^{t-1} \widetilde{D} L & RD^{t-1} L \widetilde{\mathcal{M}}_2 + \left( R \sum_{k=0}^{t-2} D^k L \right) \mathcal{M}_2 \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Since, by hypothesis,  $\mathbf{i} = (0, \dots, 0, 1, 0, \dots, 0, 0)$ , with the only 1 being at position  $i$ ,  $0 \leq i \leq N$ , then  $\left( \mathbf{i} \mathcal{M}^{t-1} \widetilde{\mathcal{M}} \right)$  denotes the  $i^{\text{th}}$ -row of matrix  $\mathcal{M}^{t-1} \widetilde{\mathcal{M}}$ . Therefore, we are interested in the sum between 0 and  $\bar{a}_s$  of the elements of the  $i^{\text{th}}$ -row of matrix  $\mathcal{M}^{t-1} \widetilde{\mathcal{M}}$ , i.e. of the  $i^{\text{th}}$ -row of matrix  $RD^{t-1} \widetilde{D} L$ , as clearly  $\bar{a}_s \leq N$ .  $D^{t-1} \widetilde{D}$  is a diagonal matrix whose  $\ell^{\text{th}}$ -diagonal element is given by:

$$\begin{aligned} \left( D^{t-1} \widetilde{D} \right)_\ell &= (2(1-r_d)r_{div}(1-r_s)\lambda_\ell + (1-r_d)(1-r_{div})(1-r_s))^{t-1} \\ &\quad \cdot (2(1-r_d)r_{div}\lambda_\ell + (1-r_d)(1-r_{div})) \\ &= (1-r_s)^{t-1} (1-r_d)^t (2\lambda_\ell r_{div} + 1 - r_{div})^t \end{aligned}$$

The result follows observing that:  $\left( RD^{t-1} \widetilde{D} L \right)_{ik} = \sum_{\ell=0}^N \left( D^{t-1} \widetilde{D} \right)_\ell r_{i\ell} l_{\ell k}$ .  $\square$

As an immediate consequence of Proposition 7, we can claim:

**Corollary 2** *For all time  $t \geq 1$  the value  $r_s(t)$  which maximizes the expected number of selected B-cells at time  $t$  is:*

$$r_s(t) = \frac{1}{t}$$

*Proof* Since  $(1-r_d)^t \sum_{\ell=0}^N (2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{\bar{a}_s} r_{i\ell} l_{\ell k}$  is a non negative quantity independent from  $r_s$ , the value of  $r_s$  which maximizes  $\mathbb{E}(S_t)$  is the one that maximizes  $r_s(1-r_s)^{t-1}$ . The result trivially follows.  $\square$

Under certain hypotheses about the mutational model and the GC evolution, one could justify the claim of Corollary 2 by heuristic arguments, without considering the  $(N+3)$ -type GW process. This leads to approximately estimate the expected number of selected B-cells at time  $t$ .

*Hypothesis 1*  $\mathcal{Q}_N$  converges through its stationary distribution, denoted by  $\mathbf{m} = (m_i)$ ,  $i \in \{0, \dots, N\}$ .

*Hypothesis 2*  $Z_t$  explodes, where  $(Z_t)_{t \in \mathbb{N}}$  is given by Definition 8.

Let  $\tilde{Z}_t$ ,  $t \geq 0$  be the random variable describing the GC-population size at time  $t$  before the selection mechanism is performed for this generation. For the sake of simplicity, let us suppose  $\tilde{Z}_0 = 1$ .  $(\tilde{Z}_t)_{t \in \mathbb{N}}$  is a MC on  $\{0, 1, 2, \dots\}$ . Denoted by  $\tilde{p}_k := \mathbb{P}(\tilde{Z}_1 = k)$ ,  $k \in \{0, 1, 2\}$ :

$$\begin{cases} \tilde{p}_0 = r_d \\ \tilde{p}_1 = (1 - r_d)(1 - r_{div}) \\ \tilde{p}_2 = (1 - r_d)r_{div} \end{cases} \quad (22)$$

It follows:  $\tilde{m} := \mathbb{E}(\tilde{Z}_1) = (1 - r_d)(1 - r_{div}) + 2(1 - r_d)r_{div} = (1 - r_d)(1 + r_{div})$ .

Conditioning to  $Z_t = k$ ,  $\tilde{Z}_{t+1}$  is distributed as the sum of  $k$  independent copies of  $\tilde{Z}_1$ , which gives:

$$\mathbb{E}(\tilde{Z}_t) = \mathbb{E}(Z_{t-1})\mathbb{E}(\tilde{Z}_1) = \mathbb{E}(Z_1)^{t-1}\mathbb{E}(\tilde{Z}_1) = (1 - r_d)^t(1 + r_{div})^t(1 - r_s)^{t-1} \quad (23)$$

Thanks to Hypotheses 1 and 2, if  $t$  is big enough, there is approximately a proportion of  $m_i$  elements in the  $i^{\text{th}}$ -affinity class with respect to  $\bar{\mathbf{x}}$ . Therefore, on average at time  $t$  there are approximately  $\sum_{i=0}^{\bar{a}_s} m_i \mathbb{E}(\tilde{Z}_t)$  B-cells in the GC belonging to an affinity class with index at most equal to  $\bar{a}_s$  with respect to  $\bar{\mathbf{x}}$ , before the selection mechanism is performed for this generation. Each one of these cells can be submitted to selection with probability  $r_s$ , and in this case it will be positively selected. Hence:

$$\mathbb{E}(S_t) \simeq r_s \sum_{i=0}^{\bar{a}_s} m_i \mathbb{E}(\tilde{Z}_t) = (1 - r_d)^t(1 + r_{div})^t(1 - r_s)^{t-1} r_s \sum_{i=0}^{\bar{a}_s} m_i, \quad (24)$$

which is maximized at time  $t \geq 1$  for  $r_s(t) = 1/t$ .

*Remark 2* One observes that the approximation in (24) gives the same value for the optimal  $r_s(t)$  as in Corollary 2. Nevertheless, it does not allow to describe exactly the behavior of  $\mathbb{E}(S_t)$ , since it is obtained by approximating the distribution of B-cells in the GC with their stationary distribution.

### 3.4 Numerical simulations

We evaluate numerically results of Proposition 6. The  $(N + 3)$ -type GW process allows a deeper understanding of the dynamics of both populations : inside the GC and in the selected pool. Through numerical simulations we emphasize the dependence of the quantities defined in Proposition 6 on parameters involved in the model.

We suppose that at the beginning of the process there is a single B-cell entering the GC belonging to the affinity class  $a_0$ . Of course, the model we set

allows to simulate any possible initial conditions. Indeed, by fixing the initial vector  $\mathbf{i}$ , we can decide to start the reaction with more B-cells, in different affinity classes. We consider  $\mathcal{Q}_N$  given by Definition 7 as transition probability matrix characterizing the mutational mechanism. When it is not stated otherwise, we set  $N = 10$ ,  $r_s = 0.1$ ,  $r_d = 0.1$ ,  $r_{div} = 0.9$ ,  $a_0 = 3$  and  $\bar{a}_s = 3$ . This parameter choice implies a small extinction probability, hence a great probability of explosion of the GC population (Corollary 1).

### 3.4.1 Evolution of the GC population

The evolution of the size of the GC can be studied by using the simple GW process defined in Section 3.1. Equation (4), in the case of a single initial B-cell, evidences that the expected number of B-cells within the GC for this model only depends on  $r_d$ ,  $r_{div}$  and  $r_s$  and it is not driven by the initial affinity, nor by the threshold chosen for positive selection  $\bar{a}_s$ , nor by the mutational rule.

Equation (4) evidences that, independently from the transition probability matrix defining the mutational mechanism, the GC size at time  $t$  increases with  $r_{div}$  and decreases for increasing  $r_s$  and  $r_d$ . Moreover, the impact of these last two parameters is the same for the growth of the GC. One could expect this behavior since the effect of both the death and the selection on a B-cell is the exit from the GC.

In order to study the evolution of the average affinity within the GC, we need to refer to the  $(N + 3)$ -type GW process defined in Section 3.2.

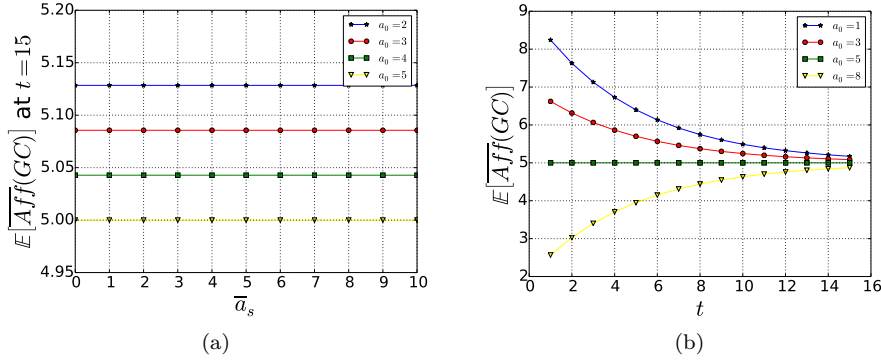


Figure 2: (a) Dependence of the expected average affinity in the GC on  $\bar{a}_s$  at time  $t = 15$ , for different values of  $a_0$ . The average affinity in the GC is constant with respect to  $\bar{a}_s$ . (b) The evolution during time of the expected average affinity in the GC for different values of  $a_0$ . The average affinity converges through  $N/2$ , due to the stationary distribution of  $\mathcal{Q}_N$ , the binomial probability distribution.

**Proposition 8** *Let us suppose that  $\mathcal{Q}_N = R\Lambda_N L$ . The average affinity within the GC at time  $t$ , starting from a single B-cell belonging to the  $i^{\text{th}}$ -affinity class with respect to  $\bar{\mathbf{x}}$  is given by:*

$$N - \frac{\sum_{\ell=0}^N (2\lambda_{\ell} r_{div} + 1 - r_{div})^t \sum_{k=0}^N k \cdot r_{i\ell} l_{\ell k}}{(1 + r_{div})^t},$$

*Proof* It follows directly from Equations (11) and (20): since we are considering the sum of the first  $N + 1$  components of the  $i^{\text{th}}$ -row of matrix  $\mathcal{M}^t$ , this is the sum of the elements of the  $i^{\text{th}}$ -row of matrix  $RD^t L$ .  $\square$

It is obvious from Proposition 8 that this quantity only depends on the initial affinity with the target trait, the transition probability matrix  $\mathcal{Q}_N$  and the division rate  $r_{div}$ . The average affinity within the GC does not depend on  $\bar{a}_s$  (as one can clearly see in Figure 2 (a)), nor by  $r_s$  or  $r_d$ . One can intuitively understand this behavior : independently from their fitness, all B-cells submitted to mutation exit the GC. Moreover,  $r_s$  and  $r_d$  impact the GC size, but not its average affinity, as selection and death affect all individuals of the GC independently from their fitness.

It can be interesting to observe the evolution of the expected average affinity within the GC during time. Simulations shows that the expected average affinity in the GC converges through  $N/2$ , independently from the affinity of the first naive B-cell (Figure 2 (b)). This depends on the mutational model we choose for these simulations. Indeed, providing that the GC is in a situation of explosion, for  $t$  big enough the distribution of GC clones within the affinity classes is governed by the stationary distribution of matrix  $\mathcal{Q}_N$ . Since for  $\mathcal{Q}_N$  given by Definition 7 one can prove that the stationary distribution over  $\{0, \dots, N\}$  is the binomial probability distribution [4], the average affinity within the GC will quickly stabilizes at a value of  $N/2$ .

### 3.4.2 Evolution of the selected pool

The evolution of the number of selected B-cells during time necessarily depends on the evolution of the GC. In particular, let us suppose we are in the supercritical case, *i.e.* the extinction probability of the GC is strictly smaller than 1. Then, with positive probability, the GC explodes and so does the selected pool. On the other hand, if the GC extinguishes, the number of selected B-cells will stabilize at a constant value, as once a B-cell is selected it can only stay unchanged in the selected pool.

As already mentioned in Section 3.3, there exists an optimal value of the parameter  $r_s$  which maximizes the expected number of selected B-cells at time  $t$ . Figure 3 (a) evidences this fact. Moreover, as expected, simulations show that the expected size of selected B-cells at a given time  $t$  increases with

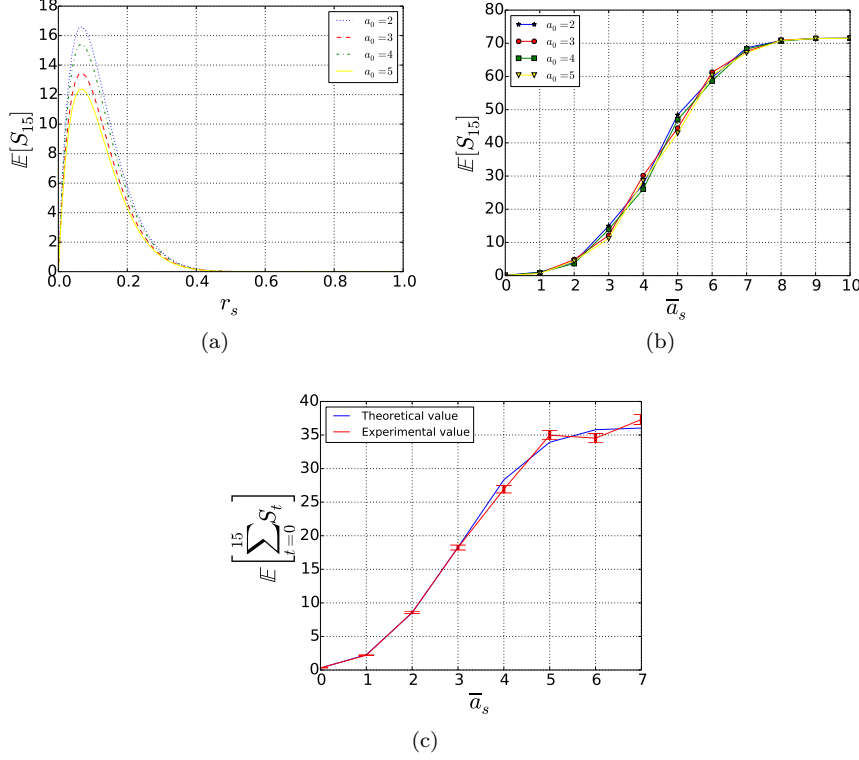


Figure 3: (a-b) Expected number of selected B-cells for the time step  $t = 15$  for different values of  $a_0$ , depending on  $r_s$  and  $\bar{a}_s$  respectively. There exists an optimal value of  $r_s$  maximizing the expected number of selected B-cells for a given time step. This value is independent from  $a_0$ . (c) Comparison between the expected number of selected B-cells until time  $t$  given by the theoretical formula (Equation (13)), and the experimental value obtained as the mean over 4000 simulations. Vertical bars denotes the corresponding estimated standard deviations. Here  $N = 7$  and  $r_s = 0.3$ .

the threshold  $\bar{a}_s$  chosen for positive selection (Figure 3 (b)). This is a consequence of Proposition 7 :  $\bar{a}_s$  determines the number of elements of the sum  $\sum_{k=0}^{\bar{a}_s} r_i \ell \ell_k$ .

Figure 3 (c) underlines the correspondence between theoretical results given by Proposition 6 and numerical values obtained by simulating the evolutionary process described by Definition 4. In particular Figure 3 (c) shows the expected (resp. average) number of selected B-cells produced until time  $t = 15$  depending on the threshold chosen for positive selection,  $\bar{a}_s$ .

## 4 Extensions of the model

Proceeding as in Section 3.2, we define and study many different models of affinity-dependent selection. Here we propose a model in which we perform only positive selection and a model reflecting a Darwinian evolutionary system, in which the selection is only negative. For the latter, we will take into account only  $N + 2$  types instead of  $N + 3$  : we do not have to consider a selected pool. Indeed the selected population remains in the GC. Here below we give the definitions of both models. In Section 4.1 we formalize these problems mathematically, then in Section 4.2 we show some numerical results.

### 4.1 Definitions and results

Let us consider the process described in Definition 4. We change only the selection mechanism.

**Definition 17 (Positive selection)** If a B-cell submitted to selection belongs to an affinity class with index greater than  $\bar{a}_s$ , nothing happens. Otherwise, the B-cell exits the GC pool and reaches the selected pool.

**Definition 18 (Negative selection)** If a B-cell submitted to selection belongs to an affinity class with index greater than  $\bar{a}_s$ , it dies. Otherwise, nothing happens.

In Figure 4 we represent schematically both processes of positive selection and of negative selection. It is clear from Figure 4 (b) that in the case of Definition 18 we do not need to consider the selected pool anymore.

#### *Positive selection*

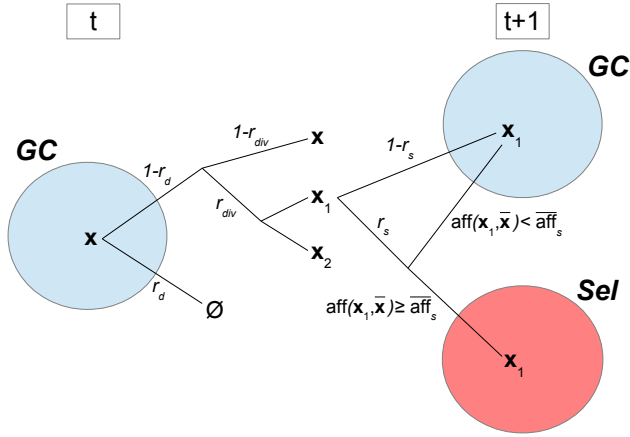
**Definition 19** Let  $\mathbf{Z}_t^{+(\mathbf{i})} = (Z_{t,0}^{+(\mathbf{i})}, \dots, Z_{t,N+2}^{+(\mathbf{i})})$ ,  $t \geq 0$  be a MC where for all  $0 \leq j \leq N$ ,  $Z_{t,j}^{+(\mathbf{i})}$  describes the number of GC B-cells belonging to the  $j^{\text{th}}$ -affinity class with respect to  $\bar{\mathbf{x}}$ ,  $Z_{t,N+1}^{+(\mathbf{i})}$  the number of selected B-cells and  $Z_{t,N+2}^{+(\mathbf{i})}$  the number of dead B-cells at generation  $t$ , when the process is initiated in state  $\mathbf{i} = (i_0, \dots, i_N, 0, 0)$ , and following the evolutionary model described by Definition 17.

Let us denote by  $\mathcal{M}^+ = (m_{ij}^+)_{0 \leq i, j \leq N+2}$  the matrix containing the expected number of type- $j$  offsprings of a type- $i$  cell corresponding to the model defined by Definition 17. We can explicitly write the value of all  $m_{ij}^+$  depending on  $r_d$ ,  $r_{div}$ ,  $r_s$ , and the elements of matrix  $\mathcal{Q}_N$ .

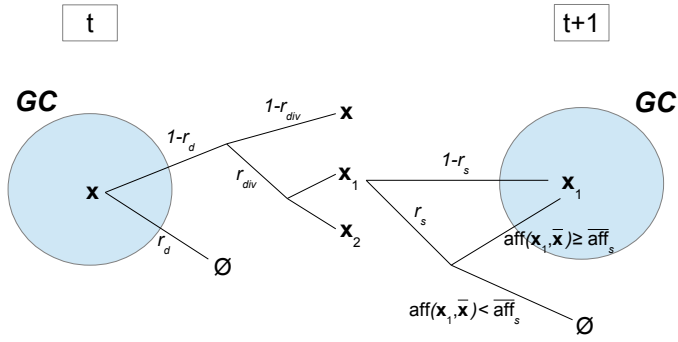
**Proposition 9**  $\mathcal{M}^+$  is a  $(N+3)^2$  matrix, which we can define as a block matrix in the following way:

$$\mathcal{M}^+ = \begin{pmatrix} \mathcal{M}_1^+ & \mathcal{M}_2^+ \\ \mathbf{0}_{2 \times (N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:



(a) Positive selection



(b) Negative selection

Figure 4: Schematic representations of models described (a) by Definitions 17 and (b) by Definitions 18 of exclusively positive (resp. exclusively negative) selection.

- $\mathcal{M}_1^+ = (m_{1,ij}^+)$  is a  $(N+1)^2$  matrix. For all  $i \in \{0, \dots, N\}$ :
  - $\forall j \leq \bar{a}_s$ :  $m_{1,ij}^+ = 2(1-r_d)r_{div}(1-r_s)q_{ij} + (1-r_d)(1-r_{div})(1-r_s)\delta_{ij}$



- $\forall j > \bar{a}_s: m_{1,ij}^+ = 2(1-r_d)r_{div}q_{ij} + (1-r_d)(1-r_{div})\delta_{ij}$   
where  $\delta_{ij}$  is the Kronecker delta.
- $\mathcal{M}_2^+ = (m_{2,ij}^+)$  is a  $(N+1) \times 2$  matrix where for all  $i \in \{0, \dots, N\}$   $m_{2,i1}^+ = m_{2,i1}$ , and  $m_{2,i2}^+ = r_d$ . We recall that  $m_{2,i1}$  is the  $i^{th}$ -component of the first column of matrix  $\mathcal{M}_2$ , given in Proposition 2.

### Negative selection

**Definition 20** Let  $\mathbf{Z}_t^{-(\mathbf{i})} = (Z_{t,0}^{-(\mathbf{i})}, \dots, Z_{t,N+1}^{-(\mathbf{i})})$ ,  $t \geq 0$  be a MC where for all  $0 \leq j \leq N$ ,  $Z_{t,j}^{-(\mathbf{i})}$  describes the number of GC B-cells belonging to the  $j^{th}$ -affinity class with respect to  $\bar{\mathbf{x}}$  and  $Z_{t,N+1}^{-(\mathbf{i})}$  the number of dead B-cells at generation  $t$ , when the process is initiated in state  $\mathbf{i} = (i_0, \dots, i_N, 0, 0)$ , and following the evolutionary model described by 18.

Let us denote by  $\mathcal{M}^- = (m_{ij}^-)_{0 \leq i,j \leq N+1}$  the matrix containing the expected number of type- $j$  offsprings of a type- $i$  cell corresponding to the model defined by Definition 20.

**Proposition 10**  $\mathcal{M}^-$  is a  $(N+2)^2$  matrix, which we can define as a block matrix in the following way:

$$\mathcal{M}^- = \begin{pmatrix} \mathcal{M}_1^- & \mathbf{m}_2^- \\ \mathbf{0}_{N+1}' & 1 \end{pmatrix}$$

Where:

- $\mathcal{M}_1^- = (m_{1,ij}^-)$  is a  $(N+1)^2$  matrix. For all  $i \in \{0, \dots, N\}$ :
  - $\forall j \leq \bar{a}_s: m_{1,ij}^- = 2(1-r_d)r_{div}q_{ij} + (1-r_d)(1-r_{div})\delta_{ij}$
  - $\forall j > \bar{a}_s: m_{1,ij}^- = 2(1-r_d)r_{div}(1-r_s)q_{ij} + (1-r_d)(1-r_{div})(1-r_s)\delta_{ij}$
- $\mathbf{m}_2^-$  is a  $(N+1)$  column vector s.t. for all  $i \in \{0, \dots, N\}$   $m_i^+ = m_{2,i2}$ ,  $m_{2,i2}$  being the  $i^{th}$ -component of the second column of matrix  $\mathcal{M}_2$ , given in Proposition 2.
- $\mathbf{0}_{N+1}'$  is a  $(N+1)$  row vector composing of zeros.

We do not prove Propositions 9 and 10, since the proofs are the same as for Proposition 2.

Results stated in Proposition 6 hold true for these new models, by simply replacing matrix  $\mathcal{M}$  with  $\mathcal{M}^+$  (resp.  $\mathcal{M}^-$ ). Of course, in the case of negative selection, as we do not consider the selected pool, we only refer to (10) and (11) quatifying the growth and average affinity of the GC. Matrix  $\mathcal{M}$  is the same for both models as only selection principles change.

Because of peculiar structures of matrices  $\mathcal{M}^+$  and  $\mathcal{M}^-$ , we are not able to compute explicitly their spectra. Henceforth we can not give an explicit formula for the extinction probability or evaluate the optimal values of the

selection rate  $r_s$  as we did in Sections 3.2 and 3.3.

Nevertheless, by using standard arguments for positive matrices, the greatest eigenvalue of both matrices  $\mathcal{M}_1^+$  and  $\mathcal{M}_1^-$  can be bounded, and hence give sufficient conditions for extinction.

**Proposition 11** *Let  $\mathbf{q}^+$  (resp.  $\mathbf{q}^-$ ) be the extinction probability of the GC for the model corresponding to matrix  $\mathcal{M}_1^+$  (resp.  $\mathcal{M}_1^-$ ).*

- If  $r_{div} \leq \frac{r_d}{1-r_d}$ , then  $\mathbf{q}^+ = \mathbf{q}^- = \mathbf{1}$ .
- If  $r_s < 1 - \frac{1}{(1-r_d)(1+r_{div})}$ , then  $\mathbf{q}^+ < \mathbf{1}$  and  $\mathbf{q}^- < \mathbf{1}$ .

*Proof* Since both matrices  $\mathcal{M}_1^+$  and  $\mathcal{M}_1^-$  are strictly positive matrices (Definition 15), the Perron Frobenius Theorem insures that the spectral radius is also the greatest eigenvalue. Then the following classical result holds [16] :

**Theorem 3** *Let  $A = (a_{ij})$  be a square nonnegative matrix with spectral radius  $\rho(A)$  and let  $r_i(A)$  denote the sum of the elements along the  $i^{th}$ -row of  $A$ . Then:*

$$\min_i r_i(A) \leq \rho(A) \leq \max_i r_i(A)$$

Simple calculations provide :

$$\begin{aligned} \min_i r_i(\mathcal{M}_1^+) &= (1-r_d)(1+r_{div}) - r_s(1-r_d) \left( 2r_{div} \min_i \sum_{j=0}^{\bar{a}_s} q_{ij} + 1 - r_{div} \right) \\ \max_i r_i(\mathcal{M}_1^+) &= (1-r_d)(1+r_{div}) - 2r_sr_{div}(1-r_d) \max_i \sum_{j=0}^{\bar{a}_s} q_{ij} \\ \min_i r_i(\mathcal{M}_1^-) &= (1-r_d)(1+r_{div}) - r_s(1-r_d) \left( 2r_{div} \min_i \sum_{j=\bar{a}_s+1}^N q_{ij} + 1 - r_{div} \right) \\ \max_i r_i(\mathcal{M}_1^-) &= (1-r_d)(1+r_{div}) - 2r_sr_{div}(1-r_d) \max_i \sum_{j=\bar{a}_s+1}^N q_{ij} \end{aligned}$$

The result follows by observing that for all  $i \in \{0, \dots, N\}$ ,  $0 \leq \sum_{j=0}^{\bar{a}_s} q_{ij}$ ,  $\sum_{j=\bar{a}_s+1}^N q_{ij} \leq 1$ , and applying Theorem 2.  $\square$

*Remark 3* One can intuitively obtain the second claim of Proposition 11, as this condition over the parameters implies that the probability of extinction of the GC for the model underlined by  $\mathcal{M}_1$  (Proposition 4) of positive and negative selection is strictly smaller than 1. Indeed keeping the same parameters for all models, the size of the GC for the model of positive and negative selection is smaller than the size of GCs corresponding to both models of only positive and only negative selection. Consequently if the GC corresponding to  $\mathcal{M}$  has a positive probability of explosion, it will be necessarily the same for  $\mathcal{M}^+$  and  $\mathcal{M}^-$ .

*Remark 4* The values of both  $\rho(\mathcal{M}_1^+)$  and  $\rho(\mathcal{M}_1^-)$  depend on  $\bar{a}_s$ , varying from a minimum of  $(1-r_d)(1+r_{div})(1-r_s)$  and a maximum of  $(1-r_d)(1+r_{div})$ . Figure 5 evidences the dependence of  $\bar{a}_s$  on the spectral radius of  $\mathcal{M}_1^+$  and  $\mathcal{M}_1^-$ , using matrix  $\mathcal{Q}_N$  given by Definition 7 as transition probability matrix.

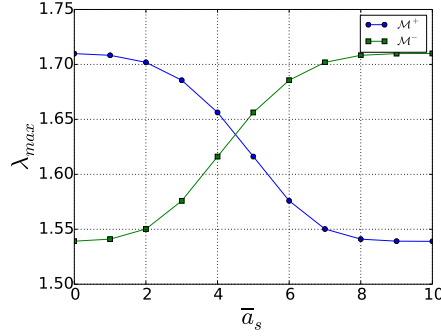


Figure 5: Dependence of greater eigenvalues of matrices  $\mathcal{M}^+$  (blue circles) and  $\mathcal{M}^-$  (green squares) respectively on  $\bar{a}_s$  for  $N = 10$ ,  $r_{div} = 0.9$ ,  $r_d = r_s = 0.1$ . Hence  $(1-r_d)(1+r_{div})(1-r_s) = 1.539$  and  $(1-r_d)(1+r_{div}) = 1.71$ .

Remark 4 and Figure 5 evidences that, conversely to the previous case of positive and negative selection, in both cases of exclusively positive (resp. exclusively negative) selection the parameter  $\bar{a}_s$  plays an important role in the GC dynamics, affecting its extinction probability. In particular, keeping unchanged all other parameters, if  $\bar{a}_s \rightarrow N$  (resp.  $\bar{a}_s \rightarrow 0$ ), then  $\rho(\mathcal{M}_1^+)$  (resp.  $\rho(\mathcal{M}_1^-)$ )  $\rightarrow (1-r_d)(1+r_{div})(1-r_s)$ , which implies  $\mathbf{q}^+$  (resp.  $\mathbf{q}^-$ )  $\rightarrow \mathbf{1}$ .

## 4.2 Numerical simulations

The evolution of GCs corresponding to matrices  $\mathcal{M}^+$  and  $\mathcal{M}^-$  respectively are complementary. Moreover, in both cases, keeping all parameter fixed one expects a faster expansion if compared to the model of positive and negative selection, since the selection acts only positively (resp. negatively) on good (resp. bad) clones. In particular, the model of negative selection corresponds to the case of 100% of recycling, meaning that all positively selected B-cells stay in the GC for further rounds of mutation, division and selection.

Figure 6 shows the dependence on  $\bar{a}_s$  of the GC size and fitness, comparing  $\mathcal{M}^+$  (left column) and  $\mathcal{M}^-$  (right column). Indeed, for these models the GC depends on the selection threshold, conversely to the previous case of positive and negative selection, and not only on the selection rate. The effects of  $\bar{a}_s$  on the GC are perfectly symmetric: it is interesting to observe that when both selection mechanisms are coupled, then  $\bar{a}_s$  does not affect the GC dynamics

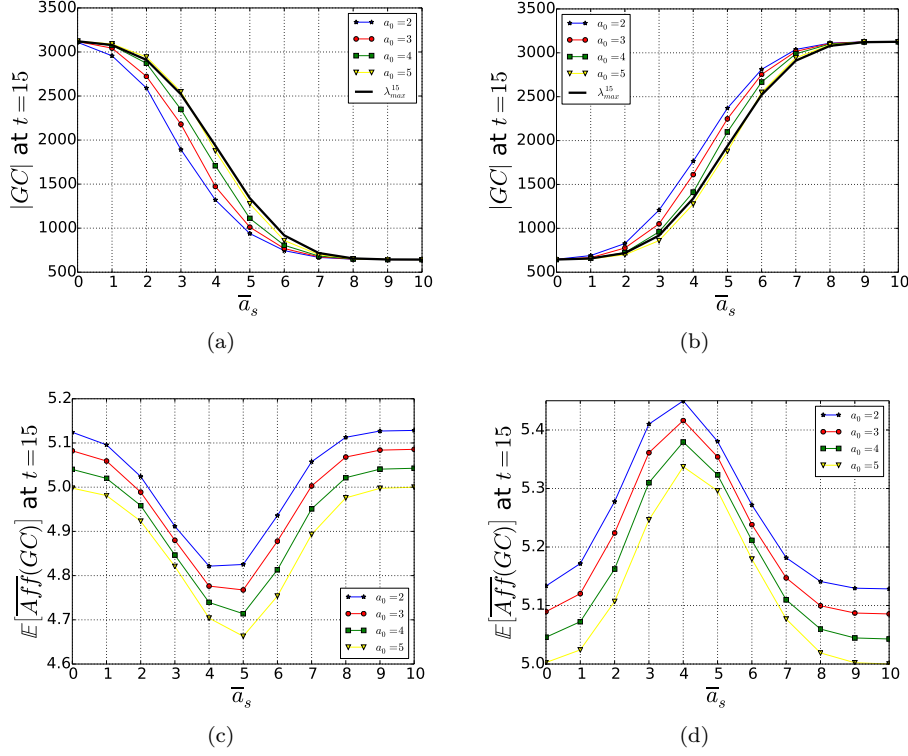


Figure 6: (a,b) Dependence of the expected size of the GC after 15 time steps on  $\bar{a}_s$  for different values of  $a_0$ . The thick black line corresponds in both figures to the value of the greater eigenvalue of matrices  $\mathcal{M}_1^+$  and  $\mathcal{M}_1^-$  respectively, raised to the power of  $t = 15$  (see Figure 5). Note that thanks to Corollary 1 we know that for this parameter choice the expected size of the GC for the model of positive and negative selection corresponds to  $((1 - r_d)(1 + r_{div})(1 - r_s))^{15}$ , which is equivalently  $\lambda_{max}^{15}$  for  $\bar{a}_s = 10$  in Figure 6(a) or  $\lambda_{max}^{15}$  for  $\bar{a}_s = 0$  in Figure 6(b). (c,d) Dependence of the expected average affinity in the GC after  $t = 15$  time steps on  $\bar{a}_s$  for different values of  $a_0$ . The left column of Figure 6 refer to the model of positive selection, while the right column to the model of negative selection.

any more, as shown for instance in Figure 2 (a). Moreover, Figures 6 (c,d) evidences the existence of a value of  $\bar{a}_s$  that minimizes (resp. maximizes) the expected average affinity in the GC for  $\mathcal{M}^+$  (resp.  $\mathcal{M}^-$ ). In both cases this value is approximately  $N/2$ . This probably depends on the transition probability matrix chosen for the mutational model, which converges to a binomial probability distribution over  $\{0, \dots, N\}$ .

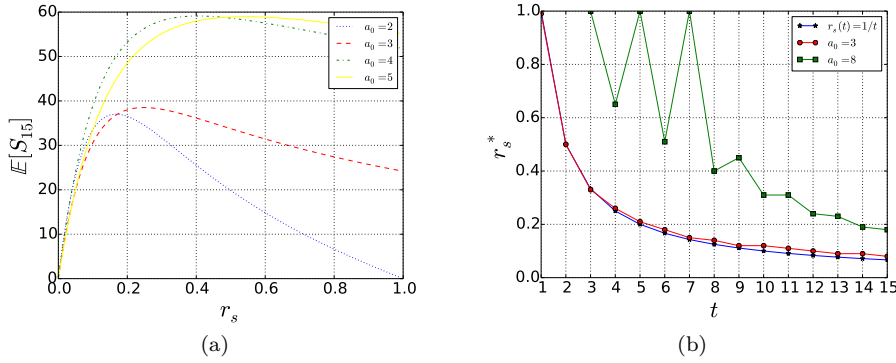


Figure 7: Model of positive selection. (a) Expected number of selected B-cells for the time step  $t = 15$  for different values of  $a_0$ , depending on  $r_s$ . (b) Estimation of the optimal  $r_s^*$  maximizing the expected number of selected B-cells for a given generation, comparing the model of positive selection for different values of  $a_0$  and the model described in Section 3 (we plot the exact value,  $r_s(t) = 1/t$ , as obtained by Corollary 2). In (b), for simulations corresponding to the model of positive selection we set  $\bar{a}_s = 5$ .

The evolution of the selected pool for the model of positive selection have some important differences if compared to the model described in Section 3. For instance, it is not easy to identify an optimal value of  $r_s$  which maximizes the expected number of selected B-cells at time  $t$ . Indeed it depends both on  $a_0$  and  $\bar{a}_s$ : if  $a_0 \leq \bar{a}_s$  we find curves similar to those plotted in Figure 3 (a), otherwise Figure 7 (a) shows a substantial different behavior. Indeed, if  $a_0 > \bar{a}_s$ , choosing a big value for  $r_s$  does not negatively affect the number of selected B-cells at time  $t$ . In this case, for the first time steps no (or a very few) B-cells will be positively selected, since they still need to improve their affinity to the target. Therefore, they stay in the GC and continue to proliferate for next generations. This fact is further underlined in Figure 7 (b), where we estimate numerically the optimal  $r_s^*$  which maximizes the expected number of selected B-cells at time  $t$ . Simulations show that for  $a_0 \leq \bar{a}_s$  the value of  $r_s^*$  for the model of positive selection is really close to the one obtained by Corollary 2. On the other hand if we start from an initial affinity class  $a_0 > \bar{a}_s$  the result we obtain is substantially different from the previous one, especially for small  $t$ . Moreover we observe important oscillations, which are probably due to the mutational model, and to the fact that the total GC size is still small for small  $t$ , since the process starts from a single B-cells. Nevertheless, it seems that for  $t$  big enough also in this case the value of  $r_s^*$  tends to approach  $1/t$ .

Since in the case of negative selection there is no selected pool, one can suppose that at a given time  $t$  the process stops and all clones in the GC pool exit the GC as selected clones. Hence it can be interesting to compare

the selected pool of the model of positive selection and the GC pool of the model of negative selection at time  $t$ . Clearly to make these two compartments comparable, the main parameters of both systems have to be opportunely chosen. In Figure 8 we compare the size and average fitness of the selected pool for  $\mathcal{M}^+$  and the GC for  $\mathcal{M}^-$  at time  $t = 30$ . We test different values of the parameter  $r_s$ . In particular, we observe that increasing  $r_s$  the GC size for the model of negative selection decreases and its average fitness increases. For the parameter choices we made for these simulations, Figure 8 (a) shows that the size of the GC for  $\mathcal{M}^-$  is comparable to the size of the selected pool for  $\mathcal{M}^+$  at time  $t = 30$  if, keeping all other parameters fixed,  $r_s = 0.15$  for  $\mathcal{M}^-$ . Nevertheless, this does not implies a comparable value for the average affinity : the clones of the selected pool for  $\mathcal{M}^+$  have a significantly greater average affinity than those of the GC for  $\mathcal{M}^-$ . In order to increase the average fitness in the GC for the model of negative selection one has to consider greater values for the parameter  $r_s$ , but this affects the probability of extinction of the process.

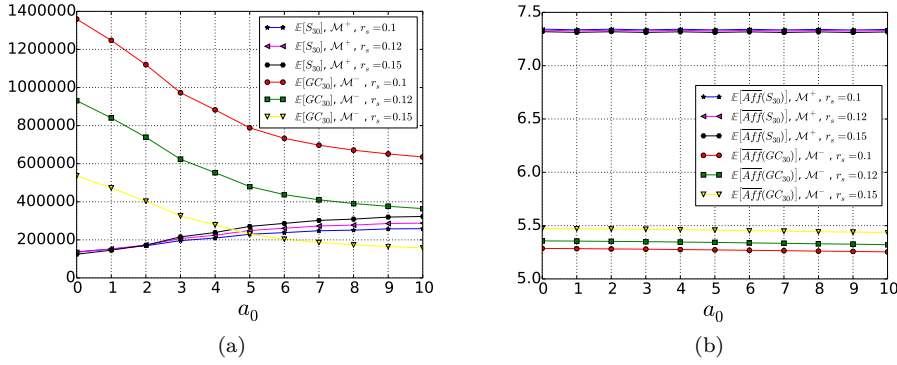


Figure 8: (a) Expected number of B-cells which have been selected until time  $t = 30$  for  $\mathcal{M}^+$  compared to the expected size of the GC for  $\mathcal{M}^-$  for different values of  $r_s$  (b) Expected corresponding average affinity for the selected pool (case of positive selection) and the GC (case of negative selection). For some choice of the parameter  $r_s$ , the size of the selected pool for  $\mathcal{M}^+$ , and the GC for  $\mathcal{M}^-$ , are comparable. Nevertheless, the corresponding average affinities are significantly different.

We can expect this discrepancy between the average affinity for the selected pool for  $\mathcal{M}^+$  and the one of the GC for  $\mathcal{M}^-$ . Indeed, in the first case we are looking to all those B-cells which have been positive selected, hence belong at least to the  $\bar{a}_s^{\text{th}}$ -affinity class. On the contrary in the case of  $\mathcal{M}^-$ , we consider the average affinity of all B-cells which are still alive in the GC at a given time step. Among these clones, if  $r_s < 1$ , with positive probability there are also

individuals with affinity smaller than the one required for escaping negative selection, which remain in the GC because they have not been submitted to selection. These B-cells make the average affinity decrease. Of course  $r_s$  is not the only parameter affecting the quantities plotted in Figure 8. In particular, one can observe that choosing a greater value for  $\bar{a}_s$  also have a significant effect over the growth of both pools, as discussed in Remark 4.

## 5 Conclusions and perspectives

In this paper we formalize and analyze a mathematical model describing an evolutionary process with affinity-dependent selection. We use a multi-type GW process, obtaining a discrete-time probabilistic model, which includes division, mutation, death and selection. In the main model developed here, we chose a selection mechanism which acts both positively and negatively on individuals submitted to selection. This leads to build matrix  $\mathcal{M}$ , which contains the expectations of each type (Proposition 2) and enables to describe the average behavior of all components of the process. Moreover, thanks to the spectral decomposition of  $\mathcal{M}$  we were able to obtain explicitly some formulas giving the expected dynamics of all types. In addition, we exhibited an optimal value of the selection rate maximizing the expected number of selected clones for the  $t^{\text{th}}$ -generation (Corollary 2).

This is one possible choice of the selection mechanism. From a mathematical point of view, the matrix  $\mathcal{M}$  is particularly easy to manipulate, as we can obtain explicitly its spectra. On the other hand, the positive and negative selection model leads, for example, to a selection threshold that does not have any impact on the evolution of the GC size. From a biological point of view this seems counterintuitive, since we could expect that the GC dynamics is sensible to the minimal fitness required for positive selection. Moreover, this process does not take into account any recycling mechanism, which has been confirmed by experiments [29] and which improves GCs' efficiency. In addition, we considered that only the selection mechanism is affinity dependent, while in the GC reaction other mechanisms, such as the death and proliferation rate, may depend on fitness [10,1]. Of course it is possible to define models with affinity-dependent division and death mechanisms with our formalism. This would clearly lead to a more complicated model, which can be at least studied numerically.

The mathematical tools used in Section 3 can be applied to define and study other selection mechanisms. For instance in Section 4 we propose two variants of the model analyzed in Section 3, in which selection acts only positively, resp. only negatively. This Section shows how our mathematical environment can be modified to describe different selection mechanisms, which can be studied at least numerically. Moreover, it gives a deeper insight of the previous model of positive and negative selection, by highlighting the effects of each selection

mechanisms individually, when they are not coupled.

From a biological viewpoint there exist many possibilities to improve the models proposed in this paper. First of all it is extremely important to fix the system parameters, which have to be consistent with the real biological process. The choice of  $N$  defines the number of affinity level with respect to a given antigen. This value can be interpreted in different ways. On the one hand it can correspond to the number of key mutations observed during the process of Antigen Affinity Maturation, hence be even smaller than 10. On the other hand, each mutational event implies a change in the B-cell affinity, slight or not if it is a key mutation. In this case the affinity can be modeled as a continuous function, hence  $N$  corresponds to a possible discretization [31, 33]. To this choice corresponds an appropriate choice of the transition probability matrix defining the mutational model over the affinity classes,  $\mathcal{Q}_N$ . In most numerical simulations we set  $N = 10$ . Nevertheless all mathematical results are independent from this choice and hold true for all  $N \geq 1$ . The selection, division and death rates have also an important impact in the GC and selected pool dynamics: in the simulations we set them in order to be in a case of explosion of the GC hence appreciate the effects of all parameters over the main quantities, but they are not biologically justified. For instance, if we suppose that a single time step corresponds to one day, then the typical proliferation rate of a B-cell has been estimated between 2 and 4 per day and in the literature we found B-cell death rates of the order of 0.5-0.8 per day [15, 34, 12].

In Section 3.3 we have explicitly determined the optimal value of the selection rate maximizing the production of output cells at time  $t$  for the main model of positive and negative selection. It is equal to  $1/t$  independently from all other parameters, and the numerical estimations we made in Section 4.2 for the model of positive selection suggested that also in this case there exists an optimal value of  $r_s(t)$ , which tends to  $1/t$  at least for  $t$  big enough. One has to interpret this result as the ideal optimal strength of the selection pressure to obtain a peak of the GC production of output cells at a given time step. For example, let us suppose that a time step corresponds to 1 day. The peak of the GC reaction has been measured to be close to day 12 [32]: for the kind of model we built and analyzed in this paper, a constant selection pressure of  $1/12$  assures a maximal production of plasma and memory B-cells coinciding with the GC peak.

In our models the selection pressure is constant. Since the optimal selection rate above depends on time, this suggests to go further in this direction. Moreover, this would allow to take into account, for instance, the early GC phase in which simple clonal expansion of B-cells with no selection occurs [9]. The hypothesis of a selection pressure changing over time can be easily integrated in our model. Indeed let us suppose that we fix a selection rate  $r_{s,1}$  until time  $t_1$  and  $r_{s,2}$  for all  $t > t_1$ . Then starting from the initial condition  $\mathbf{i}$  the expectations of each type at time  $t$  are given by  $(\mathbf{i}\mathcal{M}_{r_{s,1}}^t)$  if  $t \leq t_1$  and



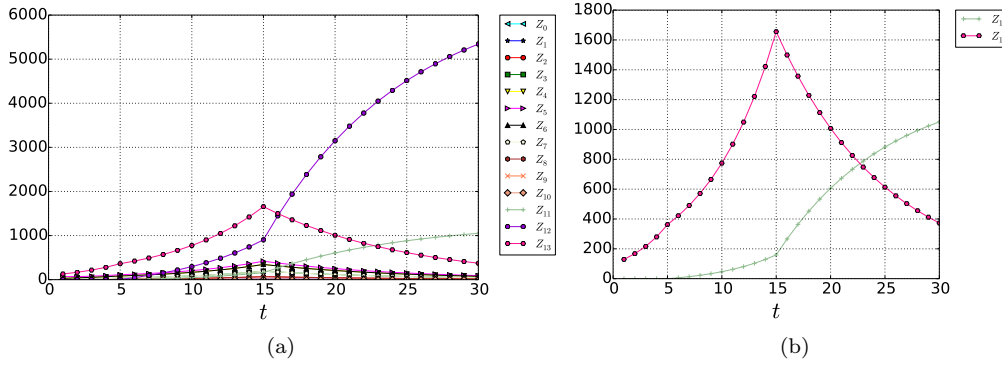


Figure 9: (a) Evolution during time of the expected value of all types for the model of positive and negative selection, with  $r_s$  varying during time and  $N = 10$ . In particular we set  $r_s = 0$  until  $t = 5$ ,  $r_s = 0.1$  from  $t = 6$  to  $t = 15$  and  $r_s = 0.3$  from  $t = 16$  to  $t = 30$ .  $Z_{13}$  denotes the total size of the GC (*i.e.*  $\sum_{k=0}^N Z_k$ ), and we recall that  $Z_{11}$  corresponds to selected B-cells and  $Z_{12}$  to dead B-cells. We set  $r_{div} = 0.3$ ,  $r_d = 0.005$  and  $z_0 = 100$  initial naive B-cells. All initial B-cells belong to  $a_0 = 5$ , and the selection threshold is  $\bar{a}_s = 3$ . (b) Evolution during time of the expected total size of the GC and the selected pool respectively, for the same set of parameters as in Figure 9 (a).

( $i\mathcal{M}_{r_{s,1}}^{t_1} \mathcal{M}_{r_{s,2}}^{t-t_1}$ ) if  $t > t_1$ , where  $\mathcal{M}_{r_{s,i}}$  is the matrix containing the expectations of each type for an evolutionary process with constant selection rate  $r_{s,i}$ ,  $i = 1, 2$ . In Figure 9 we plot the expected evolution during time of all types considering an increasing selection rate. We evaluate the expectations of all types following a process with positive and negative selection. We set  $r_s = 0$  until  $t = 5$ ,  $r_s = 0.1$  from  $t = 6$  to  $t = 15$  and  $r_s = 0.3$  for  $t > 15$ . The numerical simulations show that a time dependent selection rate allows initial explosion of the GC, and then progressive extinction, while when parameters are fixed, a GW process gives only rise either to explosion or to extinction, as shown above. The regulation and termination of the GC reaction has not yet been fully understood. In the literature, an increasing differentiation rate of the GC B-cells is thought to be a good explanation [17], here we show that other reasons could be of importance as well. Similarly, we can let other parameters vary for fixed time intervals, as well as decide to alternatively switch on and off the mutation mechanism, as already proposed in [19]. This can be obtained by alternatively use the identity matrix in place of  $\mathcal{Q}_N$ .

## References

1. Anderson, S.M., Khalil, A., Uduman, M., Hershberg, U., Louzoun, Y., Haberman, A.M., Kleinstein, S.H., Shlomchik, M.J.: Taking advantage: high-affinity b cells in the germinal

- center have lower death rates, but similar rates of division, compared to low-affinity cells. *The Journal of Immunology* **183**(11), 7314–7325 (2009)
2. Ansari, H.R., Raghava, G.P.: Identification of conformational b-cell epitopes in an antigen from its primary sequence. *Immunome research* **6**(1), 1 (2010)
  3. Athreya, K.B., Ney, P.E.: *Branching processes*, vol. 196. Springer Science & Business Media (2012)
  4. Balelli, I., Milisic, V., Wainrib, G.: Random walks on binary strings applied to the somatic hypermutation of b-cells. arXiv preprint arXiv:1501.07806 (2015)
  5. Balelli, I., Milisic, V., Wainrib, G.: Branching random walks on binary strings, with applications to adaptive immunity (2016). Preprint
  6. Bannard, O., Horton, R.M., Allen, C.D., An, J., Nagasawa, T., Cyster, J.G.: Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection. *Immunity* **39**(5), 912–924 (2013)
  7. Castro, L.N.D., Zuben, F.J.V.: Learning and optimization using the clonal selection principle. *Evolutionary Computation, IEEE Transactions on* **6**(3), 239–251 (2002)
  8. Currin, A., Swainston, N., Day, P.J., Kell, D.B.: Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews* **44**(5), 1172–1239 (2015)
  9. De Silva, N.S., Klein, U.: Dynamics of b cells in germinal centres. *Nature Reviews Immunology* **15**(3), 137–148 (2015)
  10. Gitlin, A.D., Shulman, Z., Nussenzweig, M.C.: Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature* (2014)
  11. Harris, T.E.: *The theory of branching processes*. Springer-Verlag (1963)
  12. Keşmir, C., De Boer, R.J.: A mathematical model on germinal center kinetics and termination. *The Journal of Immunology* **163**(5), 2463–2469 (1999)
  13. Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M.: Reliable b cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol* **8**(12), e1002829 (2012)
  14. MacLennan, I.C., de Vinuesa, C.G., Casamayor-Palleja, M.: B-cell memory and the persistence of antibody responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**(1395), 345–350 (2000)
  15. Meyer-Hermann, M.E., Maini, P.K., Iber, D.: An analysis of b cell selection mechanisms in germinal centers. *Mathematical Medicine and Biology* **23**(3), 255–277 (2006)
  16. Minc, H.: *Nonnegative matrices*, 1988 (1988)
  17. Moreira, J.S., Faro, J.: Modelling two possible mechanisms for the regulation of the germinal center dynamics. *The Journal of Immunology* **177**(6), 3705–3710 (2006)
  18. Pang, W., Wang, K., Wang, Y., Ou, G., Li, H., Huang, L.: Clonal selection algorithm for solving permutation optimisation problems: A case study of travelling salesman problem. In: *International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2015)*. Atlantis Press (2015)
  19. Perelson, A.S., Weisbuch, G.: *Immunology for physicists. Reviews of modern physics* **69**(4), 1219–1267 (1997)
  20. Phan, T.G., Paus, D., Chan, T.D., Turner, M.L., Nutt, S.L., Basten, A., Brink, R.: High affinity germinal center b cells are actively selected into the plasma cell compartment. *The Journal of experimental medicine* **203**(11), 2419–2424 (2006)
  21. Shannon, M., Mehr, R.: Reconciling repertoire shift with affinity maturation: the role of deleterious mutations. *The Journal of Immunology* **162**(7), 3950–3956 (1999)
  22. Shen, W.J., Wong, H.S., Xiao, Q.W., Guo, X., Smale, S.: Towards a mathematical foundation of immunology and amino acid chains. arXiv preprint arXiv:1205.6031 (2012)
  23. Shlomchik, M., Watts, P., Weigert, M., Litwin, S.: Clone: a monte-carlo computer simulation of b cell clonal expansion, somatic mutation, and antigen-driven selection. In: *Somatic Diversification of Immune Responses*, pp. 173–197. Springer (1998)
  24. Tas, J.M., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J.T., Mano, Y.M., Chen, C.S., Weill, J.C., Reynaud, C.A., Browne, E.P., et al.: Visualizing antibody affinity maturation in germinal centers. *Science* **351**(6277), 1048–1054 (2016)
  25. Timmis, J., Hone, A., Stibor, T., Clark, E.: Theoretical advances in artificial immune systems. *Theoretical Computer Science* **403**(1), 11–32 (2008)
  26. Victora, G.D.: Snapshot: the germinal center reaction. *Cell* **159**(3), 700–700 (2014)

27. Victora, G.D., Mesin, L.: Clonal and cellular dynamics in germinal centers. *Current opinion in immunology* **28**, 90–96 (2014)
28. Victora, G.D., Nussenzweig, M.C.: Germinal centers. *Annual review of immunology* **30**, 429–457 (2012)
29. Victora, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L., Nussenzweig, M.C.: Germinal center dynamics revealed by multi-photon microscopy with a photoactivatable fluorescent reporter. *Cell* **143**(4), 592–605 (2010)
30. Wang, P., Shih, C.m., Qi, H., Lan, Y.h.: A stochastic model of the germinal center integrating local antigen competition, individualistic t–b interactions, and b cell receptor signaling. *The Journal of Immunology* p. 1600411 (2016)
31. Weiser, A.A., Wittenbrink, N., Zhang, L., Schmelzer, A.I., Valai, A., Or-Guil, M.: Affinity maturation of b cells involves not only a few but a whole spectrum of relevant mutations. *International immunology* **23**(5), 345–356 (2011)
32. Wollenberg, I., Agua-Doce, A., Hernández, A., Almeida, C., Oliveira, V.G., Faro, J., Graca, L.: Regulation of the germinal center reaction by foxp3+ follicular regulatory t cells. *The Journal of Immunology* **187**(9), 4553–4560 (2011)
33. Xu, H., Schmidt, A.G., O'Donnell, T., Therkelsen, M.D., Kepler, T.B., Moody, M.A., Haynes, B.F., Liao, H.X., Harrison, S.C., Shaw, D.E.: Key mutations stabilize antigen-binding conformation during affinity maturation of a broadly neutralizing influenza antibody lineage. *Proteins: Structure, Function, and Bioinformatics* **83**(4), 771–780 (2015)
34. Zhang, J., Shakhnovich, E.I.: Optimality of mutation and selection in germinal centers. *PLoS Comput Biol* **6**(6), e1000800 (2010)